PROCESSING OF CANONICAL AND SCRAMBLED WORD ORDERS IN NATIVE AND NON-NATIVE KOREAN

BY

MYEONG HYEON KIM

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Linguistics
with a concentration in Second Language Acquisition and Teacher Education
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2019

Urbana, Illinois

Doctoral Committee:

Associate Professor Tania Ionin, Chair and Director of Research
Professor James Hye-Suk Yoon, Co-Director of Research
Adjunct Assistant Research Professor Darren Tanner
Professor Kiel Christianson

Abstract

The present study aims to investigate the nature of processing mechanisms in adult first-language (L1) and second-language (L2) sentence processing, specifically, whether adult L2 learners are capable of developing native-like sentence processing strategies. By testing a head-final language (Korean) with relatively free word order, this study also examines whether and how scrambling (switching word order between subject and object of a transitive verb) affects L1 as well as L2 sentence processing (Miyamoto & Takahashi, 2002; Yamashita, 1997). The main purpose of the present study is to examine how L1 and L2 speakers integrate linguistic input and build their sentential representations in online processing of canonical and scrambled sentences.

Previous research on L2 sentence processing disagrees as to whether and how L2 processing differs from L1 processing (e.g., Clahsen & Felser, 2006; Hopp 2006, 2010; McDonald, 2006; Sorace & Filiaci, 2006). Clahsen and Felser (2006) proposed the Shallow Structure Hypothesis (SSH), which predicts that L2 learners will rely on non-syntactic cues (e.g., plausibility or heuristic word order) instead of morphosyntactic cues (e.g., case-marking information) in sentence processing, and build less detailed syntactic representations. However, McDonald (2006) and Hopp (2006, 2010) attributed L1-L2 differences to processing difficulties rather than learners' incomplete grammatical representations (the processing capacity approach, PCA). The PCA predicts that learners with advanced (or near-native) proficiency and/or high working memory spans could perform more like native speakers. Sorace and Filiaci (2006) proposed the Interface Hypothesis (IH) to account for the non-target-like response patterns found by near-native speakers, which predicts even near-native speakers will fail to perform like native speakers if integration of syntactic knowledge with other cognitive domains (e.g., pragmatics) is required.

The present study tests the three approaches by investigating whether L2 learners of Korean use morphosyntactic (case-marking), heuristic word order, and/or non-syntactic (plausibility or information structure) information when processing scrambled sentences (i.e., the filler-gap dependencies), as compared to native Korean speakers. This study also considers whether such factors as proficiency and working memory capacities play a role in attaining native-like processing. Three experiments are implemented in the present study; the main task is an online self-paced reading task for all three, which measures participants' online comprehension of sentences. Additionally, an agent identification task tests L2 learners' offline knowledge of word order and case markers in Korean, and an offline acceptability task measures participants' knowledge of information structures in Korean. The results of the present study confirmed what the processing capacity approach (PCA) predicts, L1 and L2 differences can be attributed to processing difficulties rather than learners' incomplete grammatical representations. Thus, the findings of the study suggest that L2 processing is not qualitatively different from L1 processing.

To my parents, Youngsik Kim and Sunsim Kim

Acknowledgements

I would like to express my sincere thanks to all the people who have supported and encouraged me during this long journey. Without them, it would not have been possible to reach this far. My words cannot express my gratitude to them.

First and foremost, I am greatly thankful to my two wonderful advisors, Tania Ionin and James Hye-Suk Yoon. Throughout my years at graduate school, they have provided me with so much support, care, encouragement, and patience. Their expertise and guidance made this whole process of writing a dissertation possible and even enjoyable. They also have been great mentors to me. In the moments of disappointment, frustration, and setbacks, they always believed in me and encouraged me to keep trying. And each time I was then able to move forward without quitting. I have been so lucky to have such great people as my advisors as well as my role models.

I also want to thank the other members of the dissertation committee, Darren Tanner and Kiel Christianson, for their valuable comments on this dissertation and their great support. I have learned more about psycholinguistics and statistics through our discussions. Their detailed comments and insightful questions have helped me find better ways to analyze data and made me think more about how to do a better science. I have benefited so much from their advice and help in many aspects.

I owe a great debt of gratitude to my friends at UIUC and in Korea for their constant support and help. They have been there for me whenever I needed them. Although I cannot name them all here, I hope they know how fortunate I am to have friends like them. I would additionally like to thank to people who have been involved in this dissertation project. Heechul

Lee and Jai Park from Chonbuk National University helped me recruit participants and provided me with office space for data collection. I wish to express my deepest gratitude to the participants of this dissertation project. Their active participation in this project enabled me to write this dissertation.

Above all, I extend gratitude to my mother, my father, and my brother for their continuous support and love. I believe I am who I am today because of them. Finally, I must thank my husband, Takashi Sanehiro, who has been more than supportive throughout this whole process. He has been an anchor for me and his faith in me has been a constant source of motivation and willpower to persist and successfully complete this project. I also want to say special thanks to my daughter, Yuu Sanehiro, for her unconditional love. She has been my inspiration to continue and she has given me the strength to go forward every day.

Last but not least, I would like to thank and praise my Lord, Jesus Christ, for his grace throughout this journey. He has been with me every moment and allowed me the best in my life.

Without any of them, I could not have made it this far. Thank you for all your support and encouragement.

Table of Contents

Chapter 1

Introduction

When people comprehend utterances, the human language processor processes incoming language input incrementally in real time (e.g., Altmann & Steedman, 1988; Frazier, 1989). The linguistic input is integrated into a connected syntactic structure in a word-by-word fashion as soon as it is encountered with no need to wait for more input. Relevant syntactic and semantic constraints are applied, either immediately or sequentially depending on the theoretical model. Many psycholinguistic studies have targeted head-initial languages in attempting to account for incremental processing, and found that verb information (e.g., meaning, relevant thematic roles, or argument structures) plays an important role in building up a sentential representation in online sentence processing (Trueswell, Tanenhaus, & Kello, 1993; Boland, Tanenhaus, Garnsey, and Carlson, 1995). For example, when a person reads *Mary saw her sister…*, at the noun phrase (NP) *her sister*, the parser will be immediately committed to a syntactic analysis that processes it as an NP complement rather than the subject of a clause complement since the verb requires an NP complement. However, if the verb in the sentence is *assumed*, the opposite is expected. Although the verb allows both NP and clause complements, the NP *her sister* will be processed as the subject of a clause because the frequency of clause complements is higher for that verb.

In head-final languages, however, verb information is not available until the end of a clause. Some researchers (e.g., Pritchett, 1991) claim that head-final languages are also processed in a verb-centered manner (a head-driven model), which means that processing may not be incremental. However, questions have been raised as to whether head-final languages are processed incrementally like head-initial languages (Kamide & Mitchell, 1999; Kamide, Altmann, & Haywood, 2003; Mazuka, Itoh, & Kondo, 2002; Mitsugi & MacWhinney, 2010; Miyamoto & Takahashi, 2002; Suzuki, 2013). In recent psycholinguistic literature, incremental processing in head-final languages has been argued for Japanese (a head-final language).

Like Japanese, Korean is also a head-final language with relatively free word order compared to English (a head-initial language). In Korean grammatical relations between subjects and objects are marked by case markers, though certain word orders are deemed to be canonical, or basic. Depending on whether a sentence has canonical word order (SOV) or a scrambled order (OSV), the importance of case-marking as an information source for processing is different. Because parsing scrambled sentences cannot rely on word order, scrambled sentences must rely on case-marking information, whereas canonical word order sentences can use either word order or case marking (or both). Thus, the processing of case marking information in head-final languages can inform two relevant research questions: first, how case markers function as cues in sentence processing with regard to other types of linguistic information and second, the extent to which scrambling (non-canonical word order) affects sentence processing.

The present study extends this line of work by testing the use of case marking information in the processing of scrambled sentences with both native and L2 speakers of a head-final language, Korean. In addition to case-marking information, non-syntactic information is also manipulated to examine whether such non-syntactic information facilitates processing of sentences. By testing a head-final language (Korean) with relatively free word order, the present study allows for an examination of whether L2 learners whose native language is head-initial can shift parsing strategies from relying on rigid word order to relying on case-marking information as they integrate necessary information in real-time sentence comprehension.

One of the key issues that has received much attention in second language (L2) processing literature is whether L2 speakers can process sentences as native speakers do, and, in particular, what types of information they can utilize and prioritize in processing sentences (e.g., Dussias & Cramer Scaltz, 2008; Felser, Roberts, Gross, & Marinis, 2003; C. Jackson & Roberts, 2010; Papadopoulou & Clahsen, 2003; Havik, Roberts, van Hout, Schreuder, & Haverkort, 2009; Roberts & Felser, 2011; Marinis, Roberts, Felser, & Clahsen, 2005; Williams, 2006). By examining how L2 speakers process different word orders (canonical versus scrambled sentences) in Korean, this study aims to expand our understanding of whether adult L2 learners are capable of using the same type of information as native speakers and of developing native-like sentence processing strategies (Frenck-Mestre, 2002; Hopp, 2010; Jiang, 2004; Silva & Clahsen, 2008). Different accounts for what factors lead to differences between native and L2 processing are explored in reference to the processing of Korean (e.g., Clahsen & Felser, 2006; Hopp, 2006, 2010; McDonald, 2006; Sorace & Filiaci, 2006).

In addition to investigating L2 processing, this study seeks to contribute to the ongoing debate in the processing of head-final languages: whether or not the processing of scrambled word orders leads to processing difficulties for native speakers (Erdocia, Laka, Mestres-Missé, & Rodriguez-Fornells, 2009; Hopp, 2007; Jackson, 2008; Mazuka, Itoh, & Kondo, 2002; Miyamoto & Takahashi, 2002; Tamaoka, Sakai, Kawahara, & Miyaoka, 2003; Yamashita, 1997). The findings of the present study will shed light on how scrambling is processed by speakers of a head-final language.

The dissertation is structured as follows. In chapter 2, word order and relevant syntactic features of Korean are introduced. In addition to the linguistic background, previous empirical studies regarding L1 and L2 processing of scrambling and other relevant factors are discussed. Finally, three different approaches to account for similarities and differences between L1 and L2 sentence processing are introduced.

In chapter 3, the first experiment (Experiment 1) that was carried out to examine how native Korean speakers and L2 learners of Korean process canonical and locally scrambled word order sentences with plausible and implausible sentence meanings is reported. For L2 speakers, the extent to which individual factors such as proficiency and working memory capacities influences processing is considered. In Experiment 1, the main task was a self-paced reading task and two offline tasks, sentence recognition task and agent identification task, were also utilized to examine online and offline comprehension of sentences.

In chapter 4, the second experiment (Experiment 2) investigating long-distance (LD) scrambled sentences is presented. Experiment 2 examined the same research questions as in Experiment 1 by using a self-paced reading task as well as an offline agent identification task.

In chapter 5, the third experiment (Experiment 3), which investigated how local scrambling is processed with regard to its information structure, is presented. Information structure was manipulated based on the notions of *givenness* (i.e., previously mentioned in context) and *newness* (i.e., not mentioned in context). In addition to the information structure effect, the same research questions as in Experiments 1 and 2 were examined.

Finally, in chapter 6, results of the three experiments are summarized. In light of the three approaches outlined in chapter 2, the major findings of the three experiments are interpreted and discussed, which indicated that L1 and L2 sentence processing is not qualitatively different in using syntactic information in processing scrambled sentences. However, even those L2 learners who processed local scrambling similarly to native speakers still showed difficulties at syntax and discourse (pragmatic) interface. The implications and limitations of the study are also discussed.

Chapter 2

Background

2.1. Word order and scrambling in Korean

The target language, Korean, distinguishes grammatical relations by adding case markers to NPs (although the markers sometimes can be dropped, there are certain restrictions on case dropping). The most frequent and unmarked word order in Korean is Subject-Object-Verb (SOV), as in (1) (Shin, 2007), but scrambling produces Object-Subject-Verb (OSV) orders as well, as in (2).

(1)	Yenghi-ka	Cheli-lul	coahanta
	Yenghi-NOM	Cheli-ACC	like
(2)	Cheli-lul	Yenghi-ka	coahanta
	Cheli-ACC	Yenghi-NOM	like
	'Yenghi likes Cheli.'		

Since SOV is the unmarked order, this sentence type can be parsed using either word order or case-marking or both. However, in order to extract the intended meaning of OSV sentences, speakers or readers cannot rely on word order but must utilize case-marking information.

According to syntactic theories analyzing scrambling as movement (Mahajan, 1990; Saito, 1985; among others), two subtypes of scrambling, local and long-distance (LD) scrambling, are analyzed as follows: local scrambling as either A movement (movement from [-Case] to [+Case] like movement of the object to the subject position in a passive sentence) or A' movement (movement from [+Case] to [-Case] like *wh*-movement) but LD scrambling supposedly always as A' movement. In local scrambling, as in (2) and (3b), an element is moved within a clause, whereas in LD scrambling, as in (3c), movement does not take place within a clause but the preposed element is moved out of the clause ((3a) is a canonical word order sentence). For the rest of the paper, the filler (or moved element) is marked using a subscript *i*, and the gap position is marked with t_i.

(3)	a.	Tongswu-nun	Cinho-ka	Yenghi-lul	mannassta-ko	malhayssta
		Tongswu-TOP	Cinho-NOM	Yenghi-ACC	met-COMP	said
	b.	Tongswu-nun	Yenghi-lul$_i$	Cinho-ka t_i	mannassta-ko	malhayssta
		Tongswu-TOP	Yenghi-ACC	Cinho-NOM	met-COMP	said
	c.	Yenghi-lul$_i$	Tongswu-nun	Cinho-ka t_i	mannassta-ko	malhayssta
		Yenghi-ACC	Tongswu-TOP	Cinho-NOM	met-COMP	said
		'Tongswu said that Cinho met Yenghi.'				

Regarding local scrambling in (3b), however, there is a theoretical debate on whether local scrambling is base-generated (Bošković & Takahashi, 1998; Fanselow, 1990) or derived by movement[1] (just like long-distance scrambling) (Mahajan, 1990; Neeleman & Reinhart, 1998). According to the base-generation approach, *Yenghi-lul* in (3b) is base-generated (i.e., no overt movement) in its surface position, and it is lowered at LF (Logical Form, the level of at which a meaning of a sentence is represented) to the embedded object position in order to receive a theta role. Obligatory LF lowering account of scrambling is theoretically advantageous since this account can eliminate the "optional" nature of scrambling and contribute to a unified theory with no exception.

On the other hand, in Government and Binding theory (Saito, 1985; Hoji, 1985), it is

[1] It is also debatable whether local scrambling shows A movement properties or A' movement properties or mixed (Mahajan, 1990; Saito, 1992; Webelhuth, 1989). Since the distinction between them is not related to the main focus of this paper, this will not be further discussed.

argued that OSV scrambled sentences are derived from canonical (SOV) sentences by (overt) movement of the object NP (Yenghi-ACC) to the adjoined higher position than the subject, which leaves a trace in the object position. Under this framework, a scrambled sentence is considered to be more complex than its counterpart canonical sentence since it is created through an additional syntactic operation (i.e., movement) in which a gap is posited in the original position (i.e., the filler-gap dependency).

Although investigating the competing theories of local scrambling is not the main question of interest, the two different theories can be used to make different predictions regarding processing of scrambled sentences. On the one hand, if local scrambling is base-generated, no gap is posited, and no filler-gap dependency is expected. Thus, longer reading times at the original position of scrambled *Yenghi-lul* in (3b) are not predicted if the NP is not derived from the position. On the other hand, under the movement account, since processing scrambled sentences involves processing a filler-gap dependency, longer reading times at *mannassta-ko* (as a post-gap position) in (3b) than in (3a) are expected.

Different from local scrambling, in terms of LD scrambling, there is consensus that LD scrambling involves filler-gap dependencies. That is, it is expected to observe longer reading times at *mannassta-ko* (as a post-gap position) in (3c) than in (3a).

2.2. L1 Processing of scrambled sentences

Several studies have investigated how scrambling affects native speakers' sentence processing with different word orders in languages like Basque (e.g., Erdocia, Laka, Mestres-Missé, & Rodriguez-Fornells, 2009), Finnish (e.g., Hyönä & Hujanen, 1997), German (e.g., Hopp, 2007; Rösler, Pechmann, Streb, Röder, & Hennighausen, 1998), Japanese (e.g., Ueno & Kluender, 2003; Mazuka, Itoh, & Kondo, 2002; Nakayama, 1995; Tamaoka, Sakai, Kawahara & Miyaoka, 2003; Yamashita, 1997), Korean (e.g., Jackson, 2008), and Russian (e.g., Sekerina, 1997). Although many of the previous studies have shown that processing scrambling leads to increased processing loads, little is known about how scrambled sentences are processed and what factors affect processing of scrambled sentences in Korean. There is relevant literature on Japanese, which is similar to Korean in terms of word order and case marking, but studies on the processing of Japanese by native speakers have found conflicting results about local scrambling.

Using a self-paced reading task, Tamaoka, Sakai, Kawahara, and Miyaoka (2003) investigated whether word order (canonical (SOV) and scrambling (OSV)) and phrase length order (short-before-long and long-before-short) affected native sentence processing in Japanese. Although participants read sentences in a region-by-region format (e.g., *dansei-ga* 'man-NOM'), reading times for each phrase were combined for statistical analyses. In short-before-long with SOV condition, for example, reading times for the short phrase *wakai dansei-ga* 'a young man' (Subject), and for the long phrase *kon'iro-no zubon-o haita tyuunen-no dansei-o* 'a middle-aged man wearing blue pants' (Object), and for the verb *korosita* 'killed' were compared with the reading times of short phrases, long phrases, and verbs, respectively, for different conditions. In terms of reading times, they reported that there was no scrambling effect while a significant effect of scrambling was found in accuracy rates, measured by a sentence-correctness decision task after reading each sentence.

Yamashita (1997) also did not find scrambling effects in her self-paced reading task on Japanese, in which all target sentences had ditransitive verbs with three preceding argument NPs, and each noun was modified by an adjective as in (4) (Yamashita, 1997, p.169, 9(a)).

(4) wakai zimuin-ga mukuti-na syatyoo-ni omosiroi hon-o ageta

young secretary-NOM quiet president-DAT fun book-ACC gave

'A young secretary gave the quiet company president a fun book.' (canonical)

Reading times in four conditions – canonical, dative scrambled (NP-DAT> NP-NOM>NP-ACC), accusative scrambled (NP-ACC> NP-NOM> NP-DAT), and dative-accusative scrambled (NP-ACC> NP-DAT> NP-NOM) – were compared for each region. Yamashita (1997) reported that the reading times for each region were not significantly different across the four conditions.

In contrast, another study using a self-paced reading task by Miyamoto and Takahashi (2002) found scrambling effects in Japanese. The task included sentences with ditransitive verbs in embedded clauses in two conditions: canonical word order (NP-NOM>NP-DAT>NP-ACC> embedded ditransitive verb) and scrambled word order (NP-NOM>NP-ACC>NP-DAT>embedded ditransitive verb). The order of the dative-marked and accusative-marked NPs in the embedded clause was switched. Results showed that the third NP in scrambled sentences had longer reading times than that in canonical word order sentences.

It is particularly interesting to see how the three studies, using the same task, found different results. One possible explanation for the difference between them can be the syntactic structure of the NPs. In Yamashita (1997), each noun was modified by an adjective, which may have led to a greater processing load in computing the argument structure compared to the NPs without modifiers in Miyamoto and Takahashi (2002). In Yamashita (1997), the increased processing load for each NP may have obscured any scrambling effects in the critical region (the third NP). This hypothesized explanation is further supported by the results of another self-paced reading study by Mitsugi and MacWhinney (2010). They also created target sentences similar to those in Yamashita (1997) in terms of sentence complexity and word order (canonical, dative scrambled, accusative scrambled, and dative-accusative scrambled sentences[2]), and each noun was modified by an adjective. As in Yamashita (1997), but unlike in Miyamoto and Takahashi (2002), the results did not show any scrambling effects in native Japanese sentence processing.

In this respect, finding no scrambling effects in Tamaoka, Sakai, Kawahara, and Miyaoka (2003) could be also due to the syntactic complexity of NPs. In their SOV sentences with long-before-short structure, the long phrase *kon'iro-no zubon-o haita tyuunen-no* ***dansei-ga*** 'a middle-aged man wearing blue pants' appeared in the subject position, and the accusative marked same NP *kon'iro-no zubon-o haita tyuunen-no* ***dansei-o*** 'a middle-aged man wearing blue pants' appeared in the sentence initial position in OSV sentences with long-before-short structure. Even SOV sentences in this phrase length order condition required participants to process a relative clause before encountering the head noun of the subject NP. Since comparisons were made between phrases not words, there might be a confound due to relative clauses and that could obscure any scrambling effects. More importantly, no separate analysis for scrambling effects associated with the critical word, *dansei-o*, at the gap position was provided, and thus, it is still unclear whether there was the filler-gap dependency processing of local scrambling.

Using a different methodology, Tamaoka, Asano, Miyaoka, and Yokosawa (2014) provided different evidence for processing of locally scrambled sentences. In their eye-tracking study, they manipulated word orders using three simple NPs (without any modifiers) as arguments in sentences with ditransitive verbs as follows: canonical (S-IO-DO-V), single-

[2] In dative-accusative scrambled sentences, whereas NP-ACC preceded NP-DAT in Yamashita (1997), NP-ACC followed NP-DAT in Mitsugi and MacWhinney (2010) although they mentioned that the four word orders from Yamashita (1997) were tested.

scrambled (S-DO-IO-V), and double-scrambled (IO-DO-S-V) orders. They found longer reading times for the third NP in both single-scrambled and double-scrambled word order conditions than in the canonical word order condition, indicating scrambled effects.

Similarly, Mazuka, Itoh, and Kondo (2002) also observed increased processing load at the gap position (measured at the subject position) in their eye-tracking study with simple Object-Subject-Verb (scrambled word order) sentences compared to Subject-Object-Verb (canonical word order) sentences in Japanese. In addition to the eye-tracking study, local scrambling effects were also observed in an event-related brain potential study. Ueno and Kluender (2003) found ERP effects typical of filler-gap dependency in clause-internal local scrambling in Japanese by comparing four target sentences: in-situ demonstrative objects (no scrambling), *wh*-objects in-situ (no scrambling), scrambled demonstrative objects (yes scrambling), scrambled *wh*-objects (yes scrambling). Among the four conditions, the canonical and scrambled demonstrative conditions are presented below (see example sentences below from Ueno and Kluender (2003, p.247, (4a) and (4c))).

(5) Ano jimotono shinbun-ni yoruto

'according to the local newspaper,'

No scrambling

a. in-situ demonstrative (SOV)

sono	inochishirazuno	bokenka-ga	toto	**sore-o**	mitsuketa-ndesu-ka.
the/that	reckless	adventurer-NOM	finally	that-ACC	discovered-POL-Q

'did that reckless adventurer finally discover that?'

Yes scrambling

b. scrambled demonstrative (OSV)

sore-o sono inochishirazuno bokenka-ga toto mitsuketa-ndesu-ka.

That-ACC the/that reckless adventurer-NOM finally discovered-POL-Q

From the two conditions, the observed ERP effects included slow negative potentials between the filler (*sore-o*) and the gap (interpreted as indexing retention of a filler in working memory), and phasic LAN effects (in response to retrieving the filler from working memory) and P600 effects (related to integrating the filler-gap dependency) at the pre-gap position[3] (*bokenka-ga toto* 'adventurer-NOM finally'), and anterior negativity (indexing working memory demands on processing of scrambled sentences) at the final verb. That is, the ERP results reflected processing loads related to processing of the filler-gap dependency.

Different from local scrambling, there is a theoretical consensus that LD scrambling involves a filler-gap dependency. The findings reported in psycholinguistic studies also support the theoretical position as providing evidence of the filler-gap dependency processing of LD scrambling. For example, Hagiwara and colleagues (2007) also elicited ERP components (i.e., sustained anterior negativity and P600) indexing a filler-gap dependency in a study of LD-scrambled sentences. Likewise, Nakano, Felser, and Clahsen (2002) reported the filler-gap dependency processing of LD scrambling. They used cross-modal lexical decision experiments with Japanese native speakers. The participants listened to sentences in Japanese and were asked to complete a lexical decision task. The target words were either identical or semantically unrelated to the scrambled elements. An example of a target scrambled sentence is given in (6) (Nakano et al., 2002, p. 539).

(6) Suruto remon-o$_i$ [$_{CP/IP}$ futari-me-no hito-ga shikai-sha-ni [$_{CP}$ sono kodomo-ga

[3] As Ueno and Kluender also pointed out, the adverb position can also be considered as a post-gap position depending on where the gap is assumed to be posited (either preceding the adverb or following the adverb).

And then lemon-ACC the second person-NOM M.C.-DAT that child-NOM

onna-no hito-ni t_i nedatte-iru to] kotae-ta]

female person-DAT asking COMP] answered]

'And then, a lemon, the second person answered to the Master of Ceremonies that that child was asking the woman for.'

There were two positions where target words were presented, at a trace position and at a control position (500 ms before the trace position). The Japanese native speakers were also tested on their working memory spans. For the participants in the high memory span group, their reaction times to the identical words at the trace position were much faster than at the control position (antecedent priming effect). On the other hand, no antecedent priming effect was observed for the participants in the low memory span group. Since the sentences were long and rather complex, high working memory span seems to be required to reactivate the antecedent at the (possible) gap posited position. With the interaction between working memory span and reactivation of the antecedent, the results strongly support the claim for the filler-gap dependency processing of LD scrambling.

To sum up, there has been conflicting evidence as to whether local scrambling involves processing of the filler-gap dependency based on studies using self-paced reading tasks. Nevertheless, the literature does suggest a potential factor for the conflicting results. NP complexity (with a modifier vs. without a modifier) in scrambled sentences needs to be considered as a potential factor in order to determine scrambling effects in sentence processing. As discussed above, the syntactic complexity of NP might lead to increased processing loads in both canonical and scrambled sentences, and such processing difficulty could obscure any effects associated with word order variation.

Because the results of previous studies have suggested different answers for sentences with local scrambling, comparing local and LD scrambling in one study can help us further observe whether the two are subject to similar constraints, and hence, are likely to involve the same mechanisms (i.e., filler-gap dependency processing) in L1 sentence processing. If the two subtypes of scrambling are associated with filler-gap dependencies, similar parsing performances are expected (Miyamoto & Takahashi, 2004): longer reading times at the gap position in both local and LD scrambling conditions are expected, compared to corresponding canonical word order conditions. On the other hand, if local scrambling does not involve any movement, as claimed by some theoretical studies (e.g., Bošković & Takahashi, 1998; Fanselow, 1990; Neeleman & Reinhart, 1998), there should be no slowdown in sentences with local scrambling; there should still be slowdown in sentences with LD scrambling, which are assumed to contain a filler-gap dependency on all syntactic accounts.

2.3. Contextual information in the processing of scrambling

The information structural properties of scrambled phrases are relevant since several studies have found that discourse/contextual information could reduce the difficulty associated with the processing of scrambled sentences. For example, Kaiser and Trueswell (2004) found discourse effects facilitating the processing of scrambled sentences in their Finnish study. Although the scrambling effect was not completely eliminated, providing supportive context preceding a scrambled sentence facilitated the processing of scrambling. Similarly, in Russian (Sekerina, 2003), contextual information reduced processing cost in scrambled as well as canonical word order sentences, but scrambling still induced longer reading times compared to canonical word order sentences.

For Korean, Park (2014) examined whether native speakers showed the given-before-new preferences with regard to canonical and scrambled word order sentences in Korean. Using sentences with ditransitive verbs, an offline oral preference task was implemented to measure participants' preferences between S-IO-DO-V (canonical, subject>indirect object>direct object>verb) and S-DO-IO-V (scrambled, subject>direct object>indirect object>verb) in given-new and new-given conditions. She found that native Korean speakers preferred canonical sentences to scrambled sentences when IO was the given information, while they preferred scrambled sentences to canonical sentences when DO was the given information. In addition, there was a significant main effect of word order, such that canonical sentences were significantly more preferred than scrambled sentences in the given-new condition, and for the new-given context, scrambled sentences were less likely to be chosen than canonical sentences. That is, native speakers accepted canonical sentences for both given-new and new-given contexts, with greater preference for the given-new condition. For scrambled sentences, however, they showed much stronger preferences for the given-new context than the new-given context. She suggested that native Korean speakers showed the given-before-new preferences for both canonical and scrambled word orders, but scrambled sentences tended to be more subject to the given-before-new principle in Korean.

In an online comprehension study, Jackson (2008) tested whether the given-before-new preferences were reflected in L1 processing of local scrambling in Korean, using a whole sentence reading task.

(7) Examples from Jackson (2008, p. 88, (4a)–(4d))

Context:

Ecey phathi-eyse Chelswu-nun **chinkwu han myeng**-kwa ewulikey toyessta.

'Yesterday Chelswu happened to hang out with **a friend** at a party.'

a. Given–New, canonical

Chelswu-nun **ku chinkwu-eykey** yumyenghan miswulka-lul sokayhayssta.

Chelswu-TOP **that friend-DAT** famous artist-ACC introduced

'Chelswu introduced a famous artist to the friend.'

b. New–Given, scrambled

yumyenghan miswulka-eykey Chelswu-nun **ku chinkwu-lul** sokayhayssta.

famous artist-DAT Chelswu-TOP **that friend-ACC** introduced

'Chelswu introduced the friend to a famous artist.'

c. New–Given, canonical

Chelswu-nun yumyenghan miswulka-eykey **ku chinkwu-lul** sokayhayssta.

Chelswu-TOP famous artist-DAT **that friend-ACC** introduced

'Chelswu introduced the friend to a famous artist.'

d. Given–New, scrambled

ku chinkwu-eykey Chelswu-nun yumyenghan miswulka-lul sokayhayssta.

that friend-DAT Chelswu-TOP famous artist-ACC introduced

'Chelswu introduced a famous artist to the friend.'

As in Park (2014), there was a strong given-before-new advantage for scrambled sentences, in that native speakers read scrambled sentences in given-new condition significantly faster than in new-given condition. In contrast, there was no such information structure effect on canonical word order sentences. However, since Jackson (2008) measured only whole sentence reading times, it is still unclear whether and how the information structure affected processing of the critical word (i.e., *ku chinkwu-eykey* 'that friend-DAT') at the gap-posited (or original) position in the given-new scrambled condition.

Likewise, Koizumi and Imamura (2016) also found given-before-new advantages in their online plausibility judgment task in Japanese. When reaction times were compared, native Japanese speakers processed OSV scrambled sentences significantly faster in the given-new condition than in the new-given condition, while no such effect was found in canonical SOV sentences. Despite the differences in syntactic construction, dative sentences in Jackson (2008) and transitive sentences in Koizumi and Imamura (2016), the results from the two studies indicate that canonical word order sentences are compatible with both new-given and given-new information structures than scrambled (non-canonical) word order sentences, which are more felicitous when the order of NPs conforms to given-new construal.

In addition to analyzing the reaction times, Koizumi and Imamura (2016) also examined how information structure affects the processing of local scrambling in a self-paced reading task. Analyses of reading times at the second NP in SOV canonical and OSV scrambled sentences showed a significant interaction between information structure and word order in the participant analysis, but only marginal effect in the item analysis. In the participant analysis, there were larger reading time differences for OSV sentences between given-new and new-given contexts, compared to SOV sentences.

In sum, previous studies using an online task with local scrambling showed an effect of information structure on scrambling. The three studies discussed demonstrated significant information structure effects on scrambling regardless of the length of context sentences. In terms of how information structure was manipulated, presenting a single copular sentence as context as in Koizumi and Imamura (2016)[4] seemed sufficient to establish the relevant discourse context for the following target sentence.

These studies reveal that in order to provide a unified account of how scrambling is processed, it is important to consider the role of information structure. The present study also examines how information structure of scrambling affects the time-course of processing of scrambling, by comparing sentences with and without contextual information. By doing so, we expect to examine how native Korean speakers and L2 speakers use and integrate syntactic (case-marking information) and pragmatic information (e.g., information structure) to build their sentential representations in processing of scrambled sentences.

2.4. Sentence processing in a second language

Previous psycholinguistic research in L2 sentence processing has investigated the question of whether and how L2 processing differs from L1 processing (e.g., Clahsen & Felser, 2006; Dussias, 2001; Frenck-Mestre & Pynte, 1997; Hopp, 2006; Juffs, 2004; Marinis, Roberts, Felser, & Clahsen, 2005; Ullman, 2001). Some researchers argue that L2 learners are unable to construct detailed syntactic representations and employ different parsing strategies than native speakers (Felser, Roberts, Gross, & Marinis, 2003; Papadopoulou & Clahsen, 2003, among others). Other researchers have presented evidence against the claim that L2 processing differs fundamentally from L1 processing (Dussias, 2001; Frenck-Mestre & Pynte, 1997; Hopp, 2006; Juffs, 2004; McDonald, 2006 among others), arguing that L2 processing is not qualitatively different from L1

---

[4] For example, they used a copular sentence like *Gaimusyoo-no zikan-wa Kuroki-da* 'It is Kuroki who is the vice-minister of the Ministry of Foreign Affairs.' before a target sentence *Kuroki-ga Kaneda-o mukaeta rasii.* 'It is likely that Kuroki welcomed Kaneda.'

processing and L2 learners can acquire the same processing strategies as native speakers. A consensus cannot be found as to whether L2 processing is qualitatively different from L1 processing and it is difficult to separate the issue of L2 processing from other individual factors such as proficiency and working memory.

2.4.1. L2 processing of scrambled sentences and individual factors

Very little research has been done on how scrambled structures are processed by L2 speakers and whether it is different from the patterns found in L1 processing of the same structures. Among studies that have examined L2 processing of scrambling, three studies have found little difference between L1 and L2. Mitsugi and MacWhinney (2010) examined L2 processing of scrambling in Japanese, and claimed that L2 learners' processing is not qualitatively different from L1 processing, regardless of the learners' L1s. Although Mitsugi and MacWhinney (2010) provide a promising piece of evidence, since their conclusion is based on a null finding from both native and L2 speakers, more work is needed to determine how similar L2 processing is to L1 processing.

In Erdocia and colleague's (2014) work on Basque, data from a self-paced reading task revealed that proficient L1 Spanish L2 Basque learners showed the same reading time patterns as native Basque speakers, whose reading times were longer for OSV sentences than SOV sentences. However, in their ERP study, different ERP effects were observed at the second DP in SOV and OSV sentences, frontal P600 by L2 speakers and negativity by native Basque speakers, which suggests that learners were employing different processing strategies from native speakers. The researchers attributed the difference to a typological difference between Basque (SOV) and Spanish (SVO), but this is not definitive, as they did not include another L2 group whose native language has the same headedness as Basque.

Hopp (2006) claimed that L2 learners' proficiency also plays an important role in determining whether L2 learners are able to apply native-like processing strategies in L2 processing. Using German temporarily ambiguous SO (canonical) and OS (scrambled) word order sentences, he manipulated disambiguating cues as either case on nouns (nominative or accusative) or verbal number agreement (singular or plural) in a self-paced reading task. Other semantic or plausibility cues were controlled. In the noun case condition, case was marked on definite determiners before each noun, whereas in the verbal agreement condition, the disambiguating region was at the clause-final verb.

German native speakers and four L2 learner groups were included, advanced and near-native English L2 learners of German, and advanced and near-native Dutch L2 learners of German, in order to investigate the role of proficiency. The predictions were as follows: 1) it was predicted that participants were engaged in the default (universal) structural parsing, which prefers subject-initial interpretation. 2) When the participants encountered disambiguating syntactic cues, for example, an accusative marked NP in scrambled sentences (in the noun case condition), the parser would signal that the default structural parsing needed to be reanalyzed, which would lead to a slowdown at the first accusative marked NP (or at the verb if the disambiguating cues were given in verbal number agreement). Note that Hopp (2006) predicted that scrambling effects would be observed in the first NP, while other studies (e.g., Miyamoto & Takahashi, 2002) predicted that scrambling effects would be observed at the position where a gap (where the moved element was originally placed) was posited.

The results showed that both native speakers and the two near-native proficiency learner groups performed similarly as they showed garden path effects (slow-downs) on the basis of disambiguating syntactic cues in the absence of semantic and plausibility cues. In contrast, the two advanced learner groups, regardless of their native languages, failed to show such sensitivity to syntactic cues. Hopp (2006) concluded that depending on L2 learners' proficiency levels, L2 learners could demonstrate native-like processing patterns. Thus, the finding suggests that L2 learners' proficiency as an individual factor plays an important role in accounting for how they process scrambled sentences in their L2, consistent with results reported by other L2 studies using different constructions (e.g., C. Jackson 2008; C. Jackson & van Hell, 2011; Juffs, 2005; Juffs & Harrington, 1995).

In addition to proficiency, another individual factor that has received much attention to account for individual variability in L2 sentence processing is working memory capacity (WMC). Although L2 processing of scrambling has not been examined with WMC as a modulating variable, several studies have examined how native and L2 speakers with different WMCs process filler-gap dependencies (e.g., Dallas, DeDe, & Nicol, 2013; Dussias & Piñar, 2010; Havik, Roberts, van Hout, Schreuder, & Haverkort, 2009; Kim & Christianson, 2017; Williams, 2006). In processing filler-gap dependencies, readers have to maintain the filler until encountering the gap position while processing words between the filler and the gap. Since successful comprehension of the filler-gap constructions is closely related to the availability of cognitive resources dedicated to processing and storage, greater WMC can be advantageous for readers to successfully process filler-gap dependencies, as found in L1 processing literature (e.g., Nakano, Felser, & Clahsen, 2002; Roberts, Marinis, Felser, & Clahsen, 2007).

However, whether WMC places a role in L2 processing of filler-gap dependencies has not been answered conclusively. In an ERP study, Dallas, DeDe, and Nicol (2013) examined both WMC and proficiency as individual factors to account for L2 processing of filler-gap sentences. In their target sentences shown in (8), the filler-gap sentences were created by using *which NP*, and plausibility of sentences was manipulated depending on whether *which NP* was a plausible or implausible object of the embedded verb *threatened*.

(8) Example sentences from Dallas, DeDe, and Nicol (2013, p. 776, (3c) and (3d))

The umpire asked *which player* the coach threatened before the game. (plausible)

The umpire asked *which football* the coach threatened before the game. (implausible)

They found that native speakers demonstrated N400 effects (reflecting semantic incongruity) at the embedded verb even prior to the gap, indicating their sensitivity to the plausibility, while L2 speakers did not show any plausibility effect. However, in control sentences without filler-gap dependencies (plausible sentence: *The referee asked whether the team threatened the player before the event*. implausible sentence: *The referee asked whether the team threatened the football before the event*.), L2 learners also showed plausibility effects just like native speakers. Although L2 speakers as a group failed to integrate the meaning of the filler (*which NP*) with its verb, subsequent analyses on individual differences yielded two important findings. First, proficiency was an important predictor to account for L2 processing of the filler-gap dependency in that more proficient L2 learners were able to show sensitivity to plausibility like native speakers. On the other hand, working memory did not affect L2 native-like processing of the filler-gap sentences.

In contrast, using *wh*-question in English, Dussias and Piñar (2010) found an interaction between WMC (measured by a reading span task) and the integration of plausibility information in L2 processing of filler-gap dependencies. They manipulated the type of argument extracted (subject *wh* extraction and object *wh* extraction) and the plausibility of *wh* word as the object of the main verb as in (9) and (10) (Dussias & Piñar, 2010, p.452, (10)-(11)). That is, *who* is not a plausible object of the main verb *declare* while it can be a plausible object of the verb *know*.

(9) Who$_i$ did the police declare t_i killed the pedestrian? (-plausible & subject *wh*)

Who$_i$ did the police declare the pedestrian killed t_i? (-plausible & object *wh*)

(10) Who$_i$ did the police know t_i killed the pedestrian? (+plausible & subject *wh*)

Who$_i$ did the police know the pedestrian killed t_i? (+plausible & object *wh*)

The question was whether participants were able to use plausibility information when they had to revise their initial filler-gap dependency (i.e., *who* as the object of the main verb *declare*/*know*). For native speakers, implausible sentences were easier for them to process than plausible sentences when they were required to revise their initial parsing commitment when they encountered the word *killed* or *the*. Only L2 speakers with high WMC span showed such plausibility effects. In other words, it was only the L1 Chinese L2 English learners with high span who showed similar processing patterns with native English speakers.

It seems that L2 learners as a group may not show native-like processing of the filler-gap sentences and they may not use the same type of information as native speakers do. However, it is also important to note that individual differences have been observed in L2 processing and such individual differences raise relevant research questions about whether and to what extent individual factors, such as proficiency and WMC, contribute to accounting for L2 processing. In this respect, individual differences in L2 processing should be further explored to better understand the role of individual factors in L2 processing.

Taken together, so far, few studies have been interested in how scrambling affects L2 sentence processing in head-final languages and more work needs to be done. Unlike head-initial languages, a subcategorizing verb cannot provide any clues about the subject and object of the verb until the end of a clause. Hence, in order for sentences in head-final languages to be incrementally processed, arguments need to be processed using other types of information before the predicate is encountered (Aoshima, Yoshida, & Phillips, 2009; Kamide & Mitchell, 1999; Kamide, Altmann, & Haywood, 2003). Processing scrambled sentences requires using syntactic cues (word order and case marking information) and semantic cues in order to obtain the intended meaning of a sentence. In this respect, scrambled sentences could provide an optimal testing ground to examine whether L2 learners can make use of syntactic cues and apply syntactic processing strategies with regard to resolving the filler-gap dependency like native speakers in online sentence processing. Additionally, lexical (non-syntactic) information like plausibility can also provide evidence about what types of information L2 learners prioritize in processing of the filler-gap sentences in head-final languages. Given that L2 learners are capable of using lexical information in online sentence processing, providing both syntactic and non-syntactic information enables us to examine whether they can utilize the same type of information in the same manner as the native speakers do in parsing the filler-gap dependencies. In doing so, considering individual factors, proficiency and WMC, can give us a more detailed picture of the extent to which L2 learners can perform like native speakers in processing scrambled sentences.

Before discussing main approaches of this study, it is important to note there is a growing body of research which has provided empirical evidence on L1/L2 processing differences based on a cue-based model (see Cunnings, 2017 for review). Under the cue-based model, Cunnings proposes that differences between L1 and L2 processing are not due to L2 representational deficits nor limited processing capacities. Instead, the differences arise because L2 learners are more susceptible to retrieval interferences than native speakers during parsing, and they rely more on discourse-based cues than syntactic cues (i.e., different weighting of cues compared to native speakers) during memory retrieval. Although his proposal is interesting to connect L2 processing theory with memory-based theory, more detailed accounts are needed. Particularly, in terms of the latter claim about L2 leaners' overreliance of discourse-based cues, which is relevant to the present study, it would be important to understand why L2 learners would over-rely on discourse-based cues (or weight more on discourse-based cues) than native speakers. Otherwise, it would be difficult to make precise predictions regarding discourse-based cues (or non-syntactic cues).

2.5. Approaches to differences between L1 and L2 processing

In accounting for the differences between L1 and L2 sentence processing, Clahsen and Felser (2006) proposed the Shallow Structure Hypothesis (SSH). According to the SSH, syntactic representations that L2 learners build up are shallower (less detailed) than those of native speakers. Clahsen and Felser (2006) claim that L2 learners tend to be sensitive to semantic or pragmatic cues during parsing, but are less likely to make use of detailed syntactic information in processing sentences. Since processing scrambled sentences requires using morphosyntactic cues (case marking information) to obtain the intended meaning of a sentence, scrambling could provide an optimal testing ground for examining whether L2 learners can make use of morphosyntactic cues, apply syntactic parsing strategies, and build up abstract phrase-structures like native speakers in online processing.

Sorace and Filiaci (2006)[5] proposed the Interface Hypothesis (IH) to account for the

[5] In Sorace (2011), the hypothesis has been updated and the current version appeals to processing-based explanations for L1/L2 differences. She claims that since one of two languages needs to be inhibited (or suppressed) while activating the other, bilingualism itself places more processing load on learners' processor during online processing. Although much of studies testing the IH has been based on offline data, it is also

non-convergence found in L2 acquisition by near-native speakers. They claim that although purely syntactic properties can be fully acquired by L2 learners, structures that involve interfaces between syntax and other cognitive domains show ‘residual optionality’. That is, structures that require the integration of syntactic knowledge with that of other domains (e.g., pragmatics) are subject to incomplete acquisition, and divergence from target language grammar is found in such structures. Under this hypothesis, even though the processing of scrambling by near-native (highly proficient) speakers can be similar to native speakers, when the integration of information structure with syntactic parsing is required, even near-native speakers are predicted to fail to perform like native speakers.

McDonald (2006) and Hopp (2006, 2010) attribute L1-L2 differences in sentence comprehension to processing difficulties rather than learners’ incomplete grammatical representations. Limited cognitive resources (e.g., working memory) can affect L2 processing, resulting in less than native-like processing. That is, even though L2 learners may have the same grammatical knowledge as native speakers, when they are processing under duress, they may fail to consistently access their grammatical knowledge. This approach – referred to as the processing capacity approach (PCA) – predicts that learners with high working memory spans should perform more like native speakers.

Applied to scrambling in Korean, the SSH[6] predicts that L2 learners (regardless of

expected to observe online processing differences between native and L2 speakers at external interfaces (e.g., syntax-pragmatics) under the IH.

[6] It should be noted that even L1 sentence processing can be sometimes inaccurate and shallow (Good enough hypothesis, Christianson et al., 2001; Ferreira & Patson, 2007; Christianson, 2016). In terms of L2 sentence processing, Lim and Christianson (2013a, 2013b, 2015) have argued that good-enough hypothesis applies to both L1 and L2 sentence processing, and that less automaticity and/or capacity result in over-reliance on heuristics in L2 (and some L1) speakers. Thus, L2 processing is not qualitatively but quantitatively different from L1 processing.

proficiency and WMC) will rely on non-syntactic cues (e.g., plausibility or heuristic word order) instead of morphosyntactic cues for the processing of filler-gap dependencies, unlike native Korean speakers. On the other hand, the PCA predicts that if L2 learners have advanced (or near-native) proficiency and/or high working memory capacity, they can utilize case-marking information as native speakers in real-time sentence comprehension of scrambled sentences. To tease apart the two accounts, Experiments 1 and 2 examine how L2 learners of different proficiency levels with different working memory capacities use case marking information in the processing of scrambled sentences. Both local (Experiment 1) and LD (Experiment 2) scrambled sentences are utilized to investigate whether the two subtypes of scrambling lead to increased processing load in L2 processing, as hypothesized for L1 processing.

In addition to these two accounts, the IH is tested as well, in order to identify the locus of difficulties in L2 sentence processing. The IH predicts the failure if processing of scrambling requires the interface between syntax and other cognitive domains (e.g., pragmatics). Under this hypothesis, even near-native speakers who perform like native speakers in processing scrambled sentences will have difficulties in processing scrambled sentences with regard to information structures (pragmatics). Then, amelioration of context found in L1 processing of scrambling is not expected for L2 learners. For Experiments 1 and 2, the IH makes similar predictions to the PCA with regard to L2 processing of scrambling. Experiment 3 is included to distinguish the two approaches: the IH predicts processing difficulties (or residual optionality) at the interface between scrambling and information structure even in near-native speakers while the PCA does not necessarily predict failure due to interface difficulties.

The present study will tease apart the three approaches by using three experiments, in order to obtain a full picture of what cues (case-marking information, word order heuristic, and plausibility) are at work in L2 sentence processing of local and LD scrambling, how the processing of local scrambling is affected by informational structure profile of scrambled phrases, and how availability of specific cues is influenced by learners' proficiency and their working memory.

2.6. Research questions

Three experiments were designed to test four research questions about L1 and L2 sentence processing, described below. With regard to L1 sentence processing, RQ1, RQ2, and RQ3 are examined. With regard to L2 sentence processing, the experiments were designed to answer the overarching research question of whether adult L2 learners can make use of syntactic information when they process Korean scrambled sentences which require the processing of filler-gap dependencies. More specifically, the interaction of different cues such as word order variation, case marking and non-syntactic cues such as plausibility and information structure gives rise to the following four research questions.

RQ1. Does (local and long-distance) scrambling increase the processing load in L1 and/or L2 sentence processing?

RQ2. How do native and L2 speakers process canonical and scrambled sentences using real world plausibility as a heuristic?

2a. Do native and L2 speakers use case-marking information and/or non-syntactic cues in canonical sentences?

2b. Do native and L2 speakers use case marking information and/or non-syntactic cues in scrambled sentences?

2c. Are native and L2 speakers the same or different with regard to 2a and 2b?

RQ3. Are native and L2 speakers sensitive to information structure cues in processing canonical and scrambled sentences? If so, do information structure cues reduce the effects of scrambling in L1 and L2 sentence processing?

RQ4. Do L2 learners' proficiency and/or WMC play a role in their sentence processing? If so, is there any relationship between proficiency and working memory in L2 sentence processing?

In Experiment 1, case marking, word order, and plausibility cues are manipulated in order to determine which cues L2 learners are able to utilize. In addition, L2 learners from different proficiency levels with different working memory capacities are included to test individual differences. By doing so, Experiment 1 is designed to address research questions RQ1, RQ2, and RQ4.

In Experiment 2, sentences with long-distance scrambling are created and plausibility of the meaning of the sentences is manipulated as in Experiment 1. In addition, individual factors such as proficiency and working memory capacities are tested as well. Thus, research questions RQ1, RQ2, and RQ4 are addressed in Experiment 2.

In Experiment 3, plausibility is not manipulated (i.e., no plausibility cue is provided), which differs from Experiments 1 and 2, but instead information structure is manipulated. Same as in Experiments 1 and 2, individual factors (i.e., proficiency and WMC) are tested. Experiment 3 is designed to examine research questions RQ1, RQ3, and RQ4.

2.7. Additional background information

Before discussing experiments, it is necessary to discuss the target L2 population of this study. In this study, all L2 learners in all experiments are native speakers of Mandarin Chinese learning

Korean as L2. Their native language, like English, has SVO order (Sun & Givón, 1985), and does not have case marking. The target population is chosen in order to reduce effects of L1 (i.e., transfer effects). Since Mandarin Chinese is typologically different from Korean in word order (SVO in mandarin Chinese vs. SOV in Korean), canonical word order information from their L1 cannot provide any cues to process Korean sentences. In addition, due to the absence of the robust case system in learners' L1, their L1 does not facilitate processing of scrambled sentences in Korean which requires processing of case markers. Thus, if the L2 learners in this study can process scrambled sentences, it can be assumed that it is less likely due to L1 influences.

Chapter 3

Experiment 1: Processing of local scrambling

Experiment 1 examined how second language (L2) learners of Korean process active sentences with transitive verbs where syntactic information (case-marking) conflicts with plausibility. I manipulated the word order in sentences (scrambling) with no effect on the proposition expressed, and tried to see whether scrambling influences L2 sentence processing. This study utilized a self-paced reading task to answer the following research questions.

RQ1. Does (local and long-distance) scrambling increase the processing load in L1 and/or L2 sentence processing?

RQ2. How do native and L2 speakers process canonical and scrambled sentences using real world plausibility as a heuristic?

2a. Do native and L2 speakers use case marking information and/or non-syntactic cues in canonical sentences?

2b. Do native and L2 speakers use case marking information and/or non-syntactic cues in scrambled sentences?

2c. Are native and L2 speakers the same or different with regard to 2a and 2b?

RQ4. Do L2 learners' proficiency and/or WMC play a role in their sentence processing? If so, is there any relationship between proficiency and working memory in L2 sentence processing?

3.1. Methodology

3.1.1. Procedure (Task order)

The experiment was constructed as follows (see Figure 3.1). All the participants were given instructions in Korean for each task. First, a self-paced reading task was given to participants as a main task to measure their online comprehension of sentences in different conditions. In addition to the online task, two offline tasks were used: a sentence recognition task to make sure that participants paid attention to the meanings of sentences and an agent identification task to test L2 learners' knowledge of word order and case markers in Korean. After that, both native speakers and L2 learners took a Korean proficiency test (a multiple choice cloze test) as well as a language background questionnaire. As the last section of the study, a working memory capacity (WMC) task (Foster, Shipstead, Harrison, Hicks, Redick, & Engle, 2014) was given to only L2 learners[7]. After completing all the tasks, L2 learners but not native speakers were asked to complete a vocabulary task to check their lexical knowledge of experimental sentences from the online task.

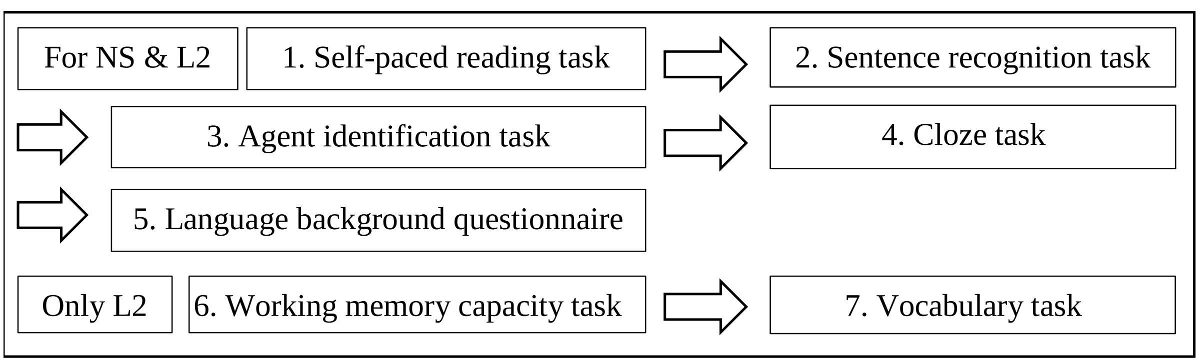

Figure 3.1. The order of tasks for Experiment 1

[7] Native speakers were not tested for their working memory capacity in Experiment 1because WMC was not main interest of study when native data was collected.

3.1.2. Participants

Twenty native Korean speakers who all resided in the United States at the time of testing participated in this study as a baseline group. All of them came to the U.S. after the age of 21. The average length of stay in the United States was about 3.3 years (max: 9 years, min: 6 months). Since the average percentage of use of Korean in their daily lives was 69% (range: 25% ~ 95%) according to their background questionnaire, it can be assumed that they were exposed to Korean in the U.S. on a regular basis.

As a target group, 40 Mandarin Chinese speaking L2 learners of Korean were tested in Korea who met the following inclusion criteria. 1) They started studying Korean after the age of 18, so their onset age of studying Korean is at least 19. It should be noted that it is not the age of their first exposure but the onset age of the first instruction, since exposure to Korean culture or language in China through media is becoming easier and earlier than before. However, early exposure to Korean culture does not necessarily mean that those exposed can speak and read Korean. Given the fact they all started studying Korean after the age of 18, the L2 learners in this study can be considered as adult second language learners. 2) They have not studied Japanese, another SOV language with similar case-marking system. 3) They studied Korean at least for 2 years, or should have a TOPIK (Test of Proficiency in Korean) certificate of level 3 or higher. TOPIK is a standardized Korean proficiency test (from level 1 (basic) to level 6 (advanced)), and level 3 is approximately intermediate level. Since a majority of L2 learners recruited in this study are college students, many of them took the test as it is required for foreign students who enter Korean universities. However, the TOPIK scores were only used for initial screening at the recruitment stage. Instead, a cloze test was utilized to determine the proficiency levels in this study since members of the target group had different language experiences with regard to both the length of formal language education and the length of stay in Korea for language learning purposes (see Table 3.1).

The Korean cloze test was developed by E-S Chung for her doctoral dissertation research (2013). The total number of blanks was 40 in this Korean multiple-choice cloze test. The reading passages were from a Korean folk tale ("Golden Axe, Silver Axe") and the story of the Three Little Pigs. The test also included items where participants were asked to find ungrammatical parts of sentences. As expected, native speakers scored at ceiling. As for L2 learners, since the scores ranged from 16 to 39, the learners are not at the same proficiency level. Thus, further analysis was conducted using proficiency (i.e., cloze scores) as a factor.

Table 3.1. Background information for the participants in Experiment 1

Group	Total number	Age	First age of instruction	Length of study Korean (Months)	Length of residence in Korea (Months)	Cloze scores (Total: 40)
NS	20 (Male: 11 Female: 9)	Mean: 28 Range: 23~33	N/A	N/A	N/A	Mean: 38.4 SD: 1 Range: 36~40
L2	40 (Male:7 Female: 33)	Mean:25 Range: 21~30	Mean: 21 Range: 19~28	Mean: 24.7 Range: 6~60	Mean: 26.8 Range: 1~120	Mean: 26.6 SD: 5.3 Range: 16~39

As the last inclusion criterion, L2 learners who scored below two standard deviation from the mean (i.e., below 16) on the cloze test was excluded. Thus, three learners who were originally tested were discarded from analysis due to not meeting the criteria. In the end, 40 L1 Mandarin Chinese L2 Korean learners were included for further analysis.

3.1.3. Materials

The self-paced reading task consisted of 48 sets of four bi-clausal Korean sentences, like (1) to (4), constructed by manipulating word order and plausibility (Ferreira, 2003; Christianson, Luke, & Ferreira, 2010; Lim & Christianson, 2013a, b; among others). Sentences within each token set started with either SOV (canonical) or OSV (scrambled) embedded clauses, and the embedded verbs were always transitive verbs requiring nominative-marked subjects and accusative-marked objects. Matrix clauses following the embedded clauses in the target items were always the same: 'NP (proper name)-NOM Verb (thought)'. A matrix clause was presented after the embedded clause to observe spill-over effects after a critical word (an embedded verb). This sentence structure is grammatically possible in Korean.

(1) Condition A: canonical, plausible

sensayngnim-i	haksayng-ul	kaluchyessta-ko	John-i	sayngkakhayssta
teacher-NOM	student-ACC	taught-COMP	John-NOM	thought

'John thought that the teacher taught the student.'

(2) Condition B: canonical, implausible

haksayng-i	sensayngnim-ul	kaluchyessta-ko	John-i	sayngkakhayssta
student-NOM	teacher-ACC	taught-COMP	John-NOM	thought

'John thought that the student taught the teacher.'

(3) Condition C: scrambled, plausible

haksayng-ul	sensayngnim-i	kaluchyessta-ko	John-i	sayngkakhayssta
student-ACC	teacher-NOM	taught-COMP	John-NOM	thought

'John thought that the teacher taught the student.'

(4) Condition D: scrambled, implausible

sensayngnim-ul	haksayng-i	kaluchyessta-ko	John-i	sayngkakhayssta
teacher-ACC	student-NOM	taught-COMP	John-NOM	thought

'John thought that the student taught the teacher.'

Two additional versions within each set were created by manipulating plausibility. *Plausibility* in this study means the ability of speakers or readers to determine if the proposition expressed by the sentence describes a typical, or plausible, state of affairs in the real world. The plausibility of a sentence containing a transitive verb was determined by whether or not the subject and object NPs of the verb are in a typical real-world relationship. For example, typically teachers teach students and not vice versa. Therefore, a sentence containing the verb *teach* will depict a plausible state of affairs if its subject denotes a teacher and its object denotes a student. It will be implausible if the referents of the subject and object are reversed. A norming task was separately administered to confirm the plausibility of target sentences and its results will be discussed in section 3.3.1.

Each condition consisted of 12 sentences, so that each participant was given 48 target sentences. Forty-one different verbs were used (7 of them were used twice). All the target sentences were presented broken down into five regions in the self-paced reading task. The critical regions were the second region (second NP), where a gap for the displaced element was posited in scrambled sentences, and the third region (embedded verb) which showed the plausible/implausible relation between the subject and the object, and the fourth region (matrix subject, a spill-over region). Experimental items were counterbalanced across four lists, so that each participant read only one version of each set. Ninety-six filler items with different syntactic structures such as subject/object relative clauses or coordinate constructions were also included. The plausibility of filler sentences was also manipulated; there were plausible, somewhat plausible, and implausible items among the fillers. Filler sentences were presented with either six or seven regions, in order to prevent participants from developing a strategy of taking the same region (i.e., third region) in all sentences as the disambiguating region. So, in total, ten practice items and 144 experimental sentences were presented (see Appendix A).

3.1.4. Self-paced reading task

The main task utilized in this study was a non-cumulative self-paced reading task (Just, Carpenter, & Woolley, 1982), given in Korean. The task was programmed and implemented using E-Prime 2.0 software. At the beginning of the task, instructions were given in Korean. In this task, participants read a series of sentences broken up into *ejeol*[8] (a basic unit of spacing in Korean, a word or a word with a case-marker), one *ejeol* at a time. At the beginning of each trial, a cross sign appeared on the left side of the screen. When participants pressed the space bar, they saw dashes corresponding to the number of syllables where each *ejeol* appeared. The dashes in the first region were replaced by an *ejeol* (or a word) when the space bar was pressed. The *ejeol* in the first region was replaced by dashes when the space bar was pressed again to reveal the next region. The end of each sentence was marked by a period. Each sentence was displayed in a single line using the uniform-width Korean font GulimChe.

After reading each sentence, participants were administered a probe recognition task where they were given a word (without a case marker) on the screen, and asked to decide whether or not the word had been used in the sentence. Half of the probe words had target

[8] Please note that 'standard' spacing conventions only approximate the word boundaries justified linguistically since *ejeol* can consist of more than one word. However, for experimental items in this study, each *ejeol* corresponds to a word such as NP-NOM, NP-ACC, Verb-COMP, or Verb.

answers of YES and corresponded to subject NPs, object NPs, or embedded verbs in the experimental sentences, while for filler sentences, the probe words could be modifiers such as adverbs. If probe words were not used in sentences and NO answers were expected, the probe words were either synonyms of words used in sentences (25%) or words that were not lexically related (25%). For example, for sentence (1) above, the probe word *kyosa* was given, with a target answer of NO. Although *kyosa* is a synonym of the word *sensayngnim* 'teacher', since it was not used in the given sentence, the target answer was NO.

The probe recognition task was used for two reasons. It was used to make sure that participants paid attention to the given sentences, and to prevent them from pressing the space bar without reading the sentences. Previous literature (e.g., Macwhinney, Bates, & Kliegl, 1984) had comprehension questions directly asking for the agents (subjects) of sentences. However, asking *who did what* type of questions could guide participants' reading in ways that the current study is interested in. If participants know what to look for before reading a sentence, the task becomes more explicit than implicit (natural reading). One possible solution to avoid such an effect is to ask about other elements in the sentences. However, since the target sentences in Experiment 1 did not have any modifiers (e.g., adjectives or adverbs), the only way to create comprehension questions would have been to ask about *who did what*. Thus, the probe recognition task was utilized instead in order to minimize the possibility of the probes leading the subjects to develop a particular reading strategy.

3.1.5. Offline Tasks

In addition to the online self-paced reading task that was expected to reveal the moment-by-moment process of participants' comprehension of each sentence, two offline tasks were used, a sentence recognition task and an agent identification task. The two offline tasks were included for different purposes. The sentence recognition task was utilized to make participants pay more attention to the meanings of the sentences in the online reading task. In the self-paced reading task instructions, participants were informed that they would be asked whether they recognized some sentences they read after completing the online task (i.e., whether the sentence was the one that they had read). They were explicitly told not to memorize the sentences but to read sentences for meaning. In order to help the participants understand the procedure, after completing the practice items in the online reading task, they were given three sentences and asked to determine whether the sentences were the ones that they had just read. In the main sentence recognition task after the online task, participants were given twelve sentences; half of them were the ones they had read in the online reading task and the other half were not. They were asked to decide if each sentence was among the ones that they had just read in the online self-paced reading task.

The second offline task, an agent identification task was used to test L2 learners' knowledge of word order and case markers in Korean. The same sentences used in the online task were included as well as half of the filler sentences. In total, three practice items and 96 sentences (48 experimental sentences and 48 filler sentences) were presented. Participants were asked to choose the subject of the embedded sentence after reading each sentence. For example, after reading the sentence (5), a question was given in Korean. Participants had to choose between the two options 1 or 2.

(5)	sensayngnim-i	haksayng-ul	kaluchyessta-ko	John-i	sayngkakhayssta
	teacher-NOM	student-ACC	taught-COMP	John-NOM	thought

'John thought that the teacher taught the student.'

Question: *John-un nuka kaluchyessta-ko sayngkakhaysssupnikka?*

‘Who did John think was teaching?’

Answer: 1. *sensayngnim* ‘teacher’

2. *haksayng* ‘student’

After completing both online and offline tasks, a cloze task and a language background questionnaire were administered to all participants, to measure Korean proficiency for L2 learners (native speakers were also tested as a baseline comparison) and to collect language background information, respectively.

3.1.6. Working Memory Capacity Tasks

For L2 learners, two additional tasks were administered, three different kinds of working memory capacity tasks and a vocabulary task. Shortened working memory tasks from Foster et al. (2014) were utilized in this study to measure L2 learners’ working memory capacities in Korean. Operation Span task (Ospan), Symmetry Span task (Sspan), and Rotation Span task (Rspan) from Foster et al. (2014) were translated from English to Korean. The basic nature of the tasks was not modified, but instructions, buttons on the screen used to answer questions, feedback sentences were translated into Korean. For Ospan task, for example, recall items were English alphabet letters, which were changed to Korean alphabet (Hangul). On the other hand, for Rspan, distractor (problem-solving) items were given in the English alphabet since they had been stored as the BMP file format (photo files) in the original task file.

All three WMC tasks, Ospan, Sspan, and Rspan, were complex span tasks, which consisted of distractor and recall items for each trial. Contrary to a simple span task, a recall item was always presented after a distractor item (see Figure 3.2). For Ospan, for example, each

Korean alphabet letter (recall item) was preceded by a simple math problem, and depending on a trial, the number of the sequences (a pair of distractor and recall items) ranged from three to seven (Figure 3.2 only has two sequences for a trial as an example). When a recall screen appeared, participants were asked to click on Korean alphabet letters in the same order that had been presented on that trial.

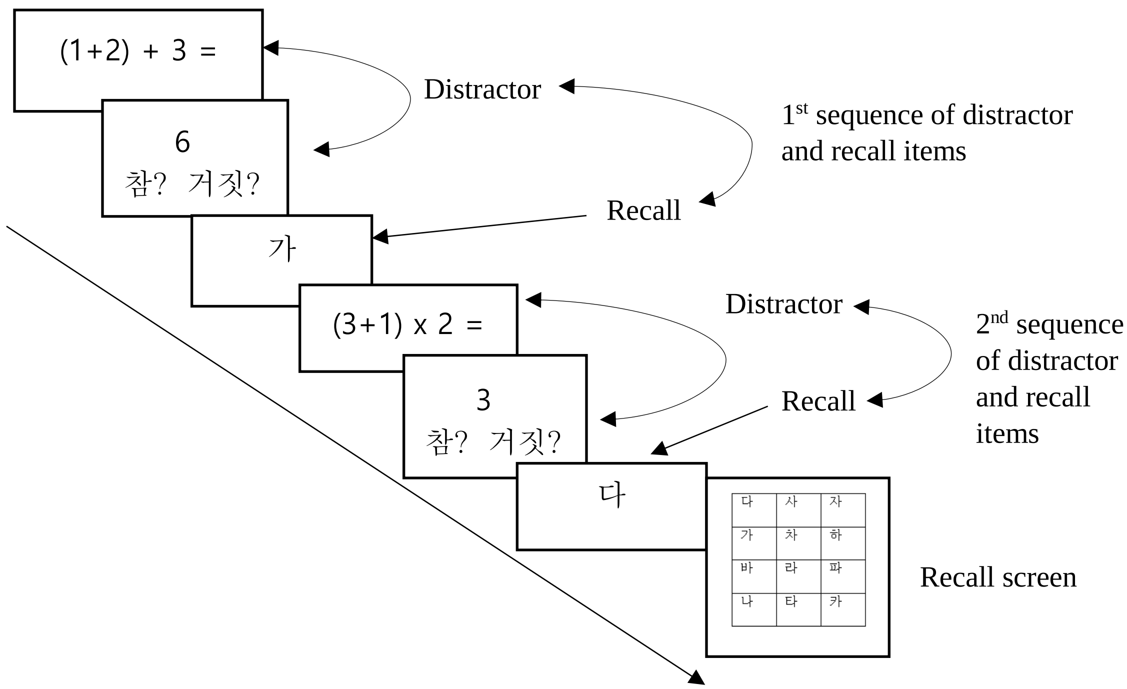

Figure 3.2. Example of operation span task

The same logic was applied to the other two WMC tasks, combining distractor and recall items for each trial. For Sspan, a distractor was a picture judgment task where participants had to judge whether a given picture was symmetrical or not, and a recall part was to remember a location of a red square. Again, each trial had different number of pairs of distractor and recall items (from two to five), and at the end of each trial, participants were asked to click the location of red squares in the same presented order. Different from the other two tasks using only Korean throughout the tasks, Rspan required participants to read English alphabet letters to judge whether a rotated English alphabet letter was left-right reversed or not[9] (distractor). After solving the rotated letter question, participants were given an arrow, and they had to remember the length (short or long) and the direction of the arrow. For each trial, the recall task was to click the arrows of the right length and direction as presented. A pair of rotated letter and arrow was presented from two to five times depending on a trial.

These shortened WMC tasks were utilized in this study for two reasons. First, since it is a shortened version of three WMC tasks, it presumably takes less amount of time than a full battery of the three WMC tasks. Despite taking less time (reducing the time by 28% of taking the full battery of the three tasks), 91% of variance of the full battery of the three WMC tasks can still be explained, according to the results from Foster et al. (2014). In addition to time advantage, the use of three WMC tasks, one verbal (Ospan) and two spatial (Sspan and Rspan) span tasks, enables us to measure L2 learners' working memory capacities more accurately without relying too much on their verbal skills. Even though the WMC tasks were given in their L2, Korean, what these WMC tasks required them to do have less to do with their language proficiency. Thus, the use of the combination of the shortened WMC tasks can be advantageous to measure L2 learners working memory capacities.

3.1.7. Vocabulary Task

After completing the three WMC tasks, L2 learners took a multiple-choice vocabulary test to measure their lexical knowledge of words used in the online self-paced reading task. 66 nouns

[9] The English letters could not be translated into Korean due to a technical reason (each English letter had been stored as a photo file). But, it should be noted that all of L2 learners in this study have experienced studying English and can read English alphabet.

and 42 verbs in Korean were tested, and four options were given in Mandarin Chinese for each question. The L2 learners were asked to choose the best answer that was most closely translated word for the Korean word.

3.2. Predictions

Based on prior literature (Kim, 1999; Miyamoto & Takahashi, 2002; among others), it is predicted that native speakers would parse sentences based on case-marking information and that if scrambling affects their reading, it would be at region 2 (second NP) or region 3 (embedded verb). Since region 2 is where a gap is posited in OSV (scrambled) sentences, a slowdown in reading time is predicted for native Korean speakers. It is also expected that scrambling effects may be observed at the next region, region 3 (a spill-over effect).

(6)	**Region 1**	**Region 2**	**Region 3**	**Region 4**	**Region 5**
	sensayngnim-i	haksayng-ul	kaluchyessta-ko	John-i	sayngkakhayssta
	teacher-NOM	student-ACC	taught-COMP	John-NOM	thought **(SOV)**
	haksayng-ul	sensayngnim-i	kaluchyessta-ko	John-i	sayngkakhayssta
	student-ACC	teacher-NOM	taught-COMP	John-NOM	thought **(OSV)**
	'John thought that the teacher taught the student.' **(Plausible)**				
	haksayng-i	sensayngnim-ul	kaluchyessta-ko	John-i	sayngkakhayssta
	student-NOM	teacher-ACC	taught-COMP	John-NOM	thought **(SOV)**
	sensayngnim-ul	haksayng-i	kaluchyessta-ko	John-i	sayngkakhayssta
	teacher-ACC	student-NOM	taught-COMP	John-NOM	thought **(OSV)**
	'John thought that the student taught the teacher.' **(Implausible)**				

Plausibility effects are predicted at region 3 (embedded verb). Native speakers would process verbs more slowly in implausible sentences than in plausible ones, regardless of word order. It is also expected that the verbs in OSV implausible sentences would be read more slowly than in other conditions due to additive effects of plausibility and spill-over effects of scrambling. Plausibility effects may also emerge on the spill-over region (region 4).

For L2 speakers, it is predicted that they would show a plausibility effect with SOV (canonical word order) sentences, not necessarily because they use case-marking to parse to obtain the correct parse. Even relying on heuristic word order (Agent> Patient>Verb) alone, L2 speakers would be able to parse SOV sentences correctly and hence show sensitivity to plausibility (because judging plausibility presupposes an accurate parse). Thus, L2 learners are expected to have longer reading times for implausible sentences than plausible sentences in SOV sentences whatever information they rely on. For OSV (scrambled) sentences, if L2 learners rely on case-marking, they are expected to be able to (i) parse these sentences although they would be slower than NSs, and (ii) show sensitivity to plausibility. If they do not use case-marking information but rely on heuristic word order and plausibility information instead, it is predicted that OSV plausible sentences would be read more slowly than OSV implausible sentences. If they do not rely on case-marking or heuristic word order exclusively but are somewhat confused with all the information, they are expected to show less determinate plausibility effects with OSV sentences in general due to failing to parse them correctly. The predictions are summarized in Table 3.2, with Prediction 1 indicating when L2 processing is not qualitatively different from L1 processing and Prediction 2 showing when they differ qualitatively.

Table 3.2. Predictions of reading time patterns at region 3 (embedded verb)

Predictions (qualitative similarity/difference)	Reading time patterns (A, B, C, and D refer to target sentence conditions)
Prediction 1. L1 processing = L2 processing	NS: A< B ≤ C < D
	L2: A< B ≤ C < D
Prediction 2. L1 processing ≠ L2 processing	NS: A< B ≤ C < D
	L2: A ≤ D < C ≤ B

A: (SOV) canonical, plausible; B: (SOV) canonical, implausible;
C: (OSV) scrambled, plausible; D: (OSV) scrambled, implausible

3.3. Results

3.3.1. Plausibility norming task

Plausibility[10] was tested using a norming task with five native Korean speakers. There were three types of subject and object relations (16 items per relation): human and human (e.g., a doctor and a patient), human and animal (e.g., a boy and a dog), and animal and animal (e.g., a cat and a mouse). For example, a sentence describing a cat chasing a mouse depicts a plausible real-world situation. The implausible version of the sentence was created by switching the subject and object NPs, as in *the mouse chased the cat*.

In the norming task, five native Korean speakers who did not participate in the main experiment were asked to rate the plausibility of the sentences on a scale of 1 to 7, with 1 as

[10] Plausibility manipulation has been widely used for L1 and L2 sentences with NPs in canonical vs. noncanonical orders (Ferreira, 2003; Christianson, Luke, & Ferreira, 2010; Lim & Christianson 2013a, b; Zhou & Christianson, 2016a, b; among others)

impossible and 7 as always possible. All the sentences were given in SOV (canonical) word order conditions, as in (1) and (2). For example, if they read the sentence *a student went to school*, since this situation is easy to imagine, they were expected to select 7 (always possible). On the other hand, if the sentence *a son gave birth to his mom* was given, since this situation is impossible, they were expected to select 1 (impossible). In addition to all experimental sentences, 76 filler items with different plausibility were also included. Instructions were given in Korean, and the task included three practice items.

Since there was the possibility that some raters might be stricter than other or have biases, ratings were normalized by converting them into z-scores. The averaged z-score for each test item was expected to factor out rater variability and eliminate potential biases in plausibility. The distribution of z-scores was plotted in Figure 3.3.

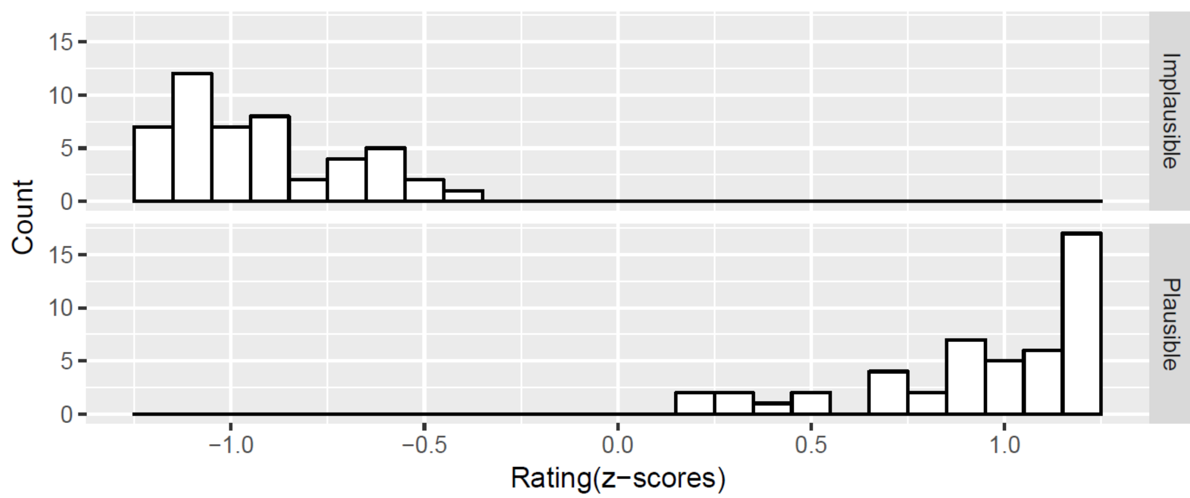

Figure 3.3. The distribution of z-scores of plausibility rating

The distribution graph above illustrates that most of plausible sentences were rated around 1 or above, whereas implausible sentences were rated around -1 or below, indicating a clear tendency that native Korean speakers accepted the plausible sentences as more plausible and the implausible sentences as less plausible.

A paired-sampled t-test was performed on the averaged z-scores for plausible and implausible sentence pairs. The difference between the plausible and implausible sentences in each set was significant: $t(47)=33.931$, $p<.001$. Significant differences were observed across the three types of plausibility relations as well. Thus, all the sentences from the norming task were retained.

3.3.2. Sentence recognition task

The average score for native speakers was 7.55 out of 12 with a standard deviation of 1.85. For L2 speakers, it was 7.4 with a standard deviation of 1.5. Since the task was added only to make participants pay attention to the sentences in the online self-paced reading task, the results of the task were not analyzed further.

3.3.3. Agent identification task

For native speakers, when they were asked to select the subject/agent of each sentence, the accuracy rate for target items was 98% (944 out of 960), and the rate including filler items was 98.8%. Twelve out of 20 native speakers correctly answered all the questions of the target items (see descriptive statistics results in Table 3.3).

For L2 speakers, the overall accuracy rate was 90.6% (3480 out of 3840) including filler items. The accuracy rate for the target items was 86.3% (1657 out of 1920). Since the filler items were sentences with subject or object relative clauses (i.e., some sentences had two nominative marked nouns), it was unlikely that L2 speakers guessed the answers without reading the sentences. The high accuracy rate for the L2 speakers in this task showed that they could comprehend both canonical and scrambled sentences in Korean.

Table 3.3. Descriptive statistics results of agent identification task

Group	Word Order	Plausibility			
		Plausible		Implausible	
		Proportion Correct	SD	Proportion Correct	SD
NS	Canonical	0.996	0.0186	1.00	0
	Scrambled	0.975	0.0476	0.962	0.0572
L2	Canonical	0.875	0.190	0.835	0.224
	Scrambled	0.919	0.117	0.823	0.196

Analyses using a generalized linear mixed model with a logit link (Barr, Levy, Scheepers, & Tily, 2013) were conducted on the target items, in R (version 3.4.3; R Core Team, 2017) with *lme4* (Bates, Maechler, & Bolker, 2015) for modeling and *lmerTest* package for significance testing (i.e., to get *p*-values).

NS and L2 data were separately analyzed. A continuous predictor, cloze scores, for L2 data were mean-centered, and categorical predictors, word order and plausibility, for both NS and L2 data were contrast coded (-0.5 vs. 0.5) before fitting models. As fixed effects, plausibility and word order were included in both models for NS and L2 data, and cloze scores only for the L2 model. Subjects and items were treated as crossed random effects. Initially, all models contained maximal random effects structures but the models failed to converge. Performing a backward model selection approach, models that would converge were identified with maximal random effects structures supported by the data. The NS fitted model included an uncorrelated by-subject random intercept and by-subject random slope of word order and an uncorrelated by-item random intercept and by-item random slope of plausibility. The L2 final model included a by-subject random intercept and a by-subject random slope of plausibility as well as a by-item random intercept. A full summary of the statistical analyses can be found in Appendix B.

For native speakers, the results showed that there was no significant main effect of plausibility (estimate = -9.359, SE= 62.922, z= -0.149, p= 0.882), and no significant main effect of word order (estimate = -11.542, SE= 62.961, z= -0.183, p= 0.855), and no interaction between them (estimate = 17.496, SE= 125.814, z= 0.139, p= 0.889). These results indicate that word order and plausibility did not significantly affect the accuracy rates of native speakers on the offline agent identification task.

For L2 speakers, different results were observed, showing significant main effects of plausibility (estimate = 0.707, SE= 0.248, z= 2.853, p= 0.004) and cloze scores (estimate = 0.115, SE= 0.046, z= 2.518, p= 0.012). Although no significant main effect of word order was found (estimate = 0.262, SE= 0.158, z= 1.657, p=0.098), a two-way interaction between plausibility and word order (estimate =0.674, SE=0.316, z=2.131, p=0.033) was present.

These results indicate that L2 learners scored better on plausible sentences than on implausible ones, and proficiency also affected their accuracy. More proficient L2 learners tended to score better. The significant interaction effect between plausibility and word order was plotted using *emmeans* package, version 1.1.2 (Lenth, 2018) via *emmip* function (see Figure 3.4 for interaction plot) in order to interpret the relationship.

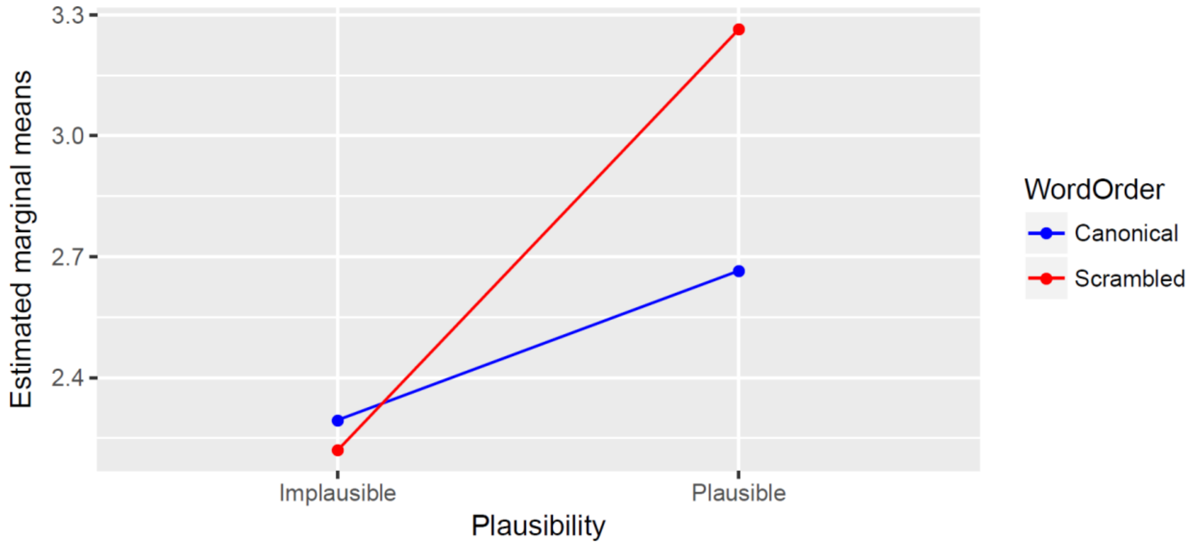

Figure 3.4. L2 interaction plot between plausibility and word order on the agent identification task. Note that the values on the y-axis are estimated marginal means of accuracy on the logit scale from the fitted L2 model.

It looks like there was really no difference in the implausible condition. It is in the plausible condition that there was a difference. Although both accuracies obviously increased, if sentential meanings were consistent with learners' world knowledge, they could interpret scrambled sentences accurately, even more accurately than canonical sentences.

3.3.4. Working Memory Capacities

Only L2 learners were tested using three shortened WMC tasks (Operation Span task (Ospan), Symmetry Span task (Sspan), and Rotation Span task (Rspan)). Two scores can be obtained for each WMC task: accuracy rates on distractor items and span scores on recall items. For accuracy rates on the distractor items, the original study (Foster et al., 2014) instructed participants to keep the accuracy rate at least at 85%. Although the same instruction was given to the participants in this study to make them pay attention to the tasks, this 85% cut-off rule was not applied since the distractor component is not the main focus of this study. Moreover, the tasks were shortened versions (i.e., using one block out of three blocks), so it would be too strict to drop participants based on 85% accuracy rate.

The average accuracy rate for Ospan on the distractor task (i.e., solving math problems) was 21.45 out of 25 (85.8%) with the standard deviation of 3.52. With shorter length than Ospan, both Sspan and Rspan had the same total number of trials, 14. The average accuracy rate for Sspan on the distractor task, judging a symmetrical picture, was 12.18 (87%) with the standard deviation of 2.33. Although Rspan could be the hardest since the L2 learners had to judge rotated English alphabet letters, the average accuracy rate was 12.03 (85.9%) with the standard deviation of 1.70, which is almost the same as the accuracy rate of Ospan.

Regarding span scores, there are different ways to score them (see Conway, Kane, Bunting, Hambrick, Wilhelm, & Engle, 2005 for more discussion about scoring). Among them, two span scores, absolute span scores and partial span scores, were obtained from the tasks. Absolute span scores are computed when participants correctly recalled items without any error, while partial span scores are calculated as the sum of all the items in the correct order within the trial. For example, if a participant correctly recalled three out of four items in the correct order, her absolute span score on this trial would be 0 (because of an error), while her partial span score would be three. Given the main interest of this study, partial span scores seem to be a better option to measure and discriminate working memory capacities between participants. Thus, the sum of partial span scores from three WMC tasks were used for further analysis in this study as WMC scores. The descriptive statistics of WMC scores for L2 learners (i.e., partial span scores) for each WMC task is presented in Table 3.4.

Table 3.4. WMC scores of three WMC tasks (L2 learners)

Ospan (total: 25)	Sspan (total: 14)	Rspan (total: 14)	WMC (total: 53)
Average: 12.7	Average: 9.8	Average: 8.6	Average: 31.2
SD: 4.8	SD: 2.8	SD: 3.0	SD: 7.6
Range: 3~22	Range: 3~14	Range: 0~13	Range: 9~46

3.3.5. Vocabulary task

Before discussing online results, it is important to make sure if L2 learners understand the meanings of the words used in the self-paced reading task. The average score of 40 L2 participants on the vocabulary test was 98.2 out of 108 (91%) with the standard deviation of 5.9. The minimum score was 81 and the maximum score was 108. Since the average score was more than 90% of the total number of scores (only one participant scored less than 80%), it seems reasonable to consider that the L2 participants in this study know the meanings of most words used in the online task. Thus, no items were excluded for further analysis based on the vocabulary task results.

3.3.6. Self-paced reading task

3.3.6.1. Probe recognition task

The mean accuracy rates of the probe recognition task were high for both native speakers (98.5% - overall, 98.6% - target items) and L2 speakers (93.1% - overall, 92% - target items) (see Table 3.5 for descriptive statistics results).

Table 3.5. Descriptive statistics results of online probe recognition task

Group	Word Order	Plausibility			
		Plausible		Implausible	
		Proportion Correct	SD	Proportion Correct	SD
NS	Canonical	0.992	0.026	0.983	0.034
	Scrambled	0.983	0.034	0.988	0.031
L2	Canonical	0.942	0.098	0.921	0.102
	Scrambled	0.908	0.123	0.910	0.117

Between SOV (canonical) and OSV (scrambled) sentences, the accuracy difference was small for both native speakers (98.8%- SOV, 98.5% - OSV) and L2 speakers (93.1% - SOV, 90.9% - OSV). Similarly, plausibility did not affect accuracy much for either group (NS: 98.5% - Implausible, 98.8% - Plausible; L2: 91.6% - Implausible, 92.5% - Plausible).

In order to analyze the online accuracy data on the target items, a generalized linear mixed model with a logit link (Barr, Levy, Scheepers, & Tily, 2013) was performed, in R (version 3.4.3; R Core Team, 2017) with *lme4* (Bates, Maechler, & Bolker, 2015) for modeling and *lmerTest* package to get *p*-values.

As with the offline comprehension data analysis, NS and L2 data were separately analyzed, and plausibility and word order were included as fixed effects for both native and L2 models, and cloze test scores only for the L2 model. Initial NS and L2 models contained maximal random effects structures but they failed to converge. A backward model selection approach was used to identify a fitted model with maximal random effects structures justified by the data that would converge. Before fitting the models, a continuous predictor, cloze scores, was centered, and categorical variables, word order and plausibility, were contrast coded with -0.5 and 0.5. The NS final model included an uncorrelated by-subject random intercept and by-subject random slope of word order and an uncorrelated by-item random intercept and by-item random slope of word order. The L2 fitted model without correlations between random effects included a by-subject random intercept and by-subject random slopes of plausibility and word order and a by-item random intercept and by-item random slopes of plausibility and word order and cloze scores. A full summary of the statistical analyses can be found in Appendix B.

Separate analyses for NS and L2 data showed no significant main effects of plausibility (NS: estimate = 0.218, SE= 0.591, z= 0.368, p= 0.713, L2: estimate =0.622, SE=0.322, z=1.932, p=0.053), word order (NS: estimate = -0.199, SE= 1.743, z= -0.114, p= 0.909, L2: estimate =-0.206, SE=0.298, z=-0.690, p=0.490), and cloze scores (L2: estimate =0.043, SE=0.039, z=1.110, p=0.267). For native speakers, no interaction was found between plausibility and word order (estimate =-0.987, SE=1.184, z=-0.834, p=0.404). For L2 speakers, however, there was a three-way interaction among plausibility, word order, and cloze scores (estimate =0.177, SE=0.081, z=2.191, p=0.028). An interaction graph was built to look closely at the relationship between the three variables, using *emmeans* package, version 1.1.2 (Lenth, 2018) via *emmip* function as below (Figure 3.5).

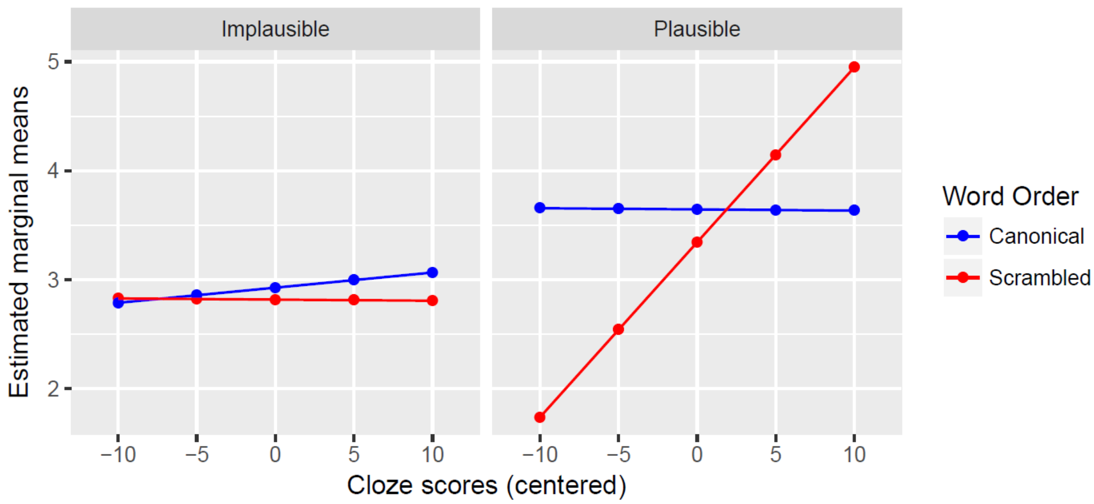

Figure 3.5. L2 interaction plot among plausibility, word order, and cloze scores on the probe recognition task. Note that cloze scores on the x-axis are mean-centered (0 equals the average) and the values on the y-axis are estimated marginal means of accuracy on the logit scale from the fitted L2 model.

This graph shows that there was nearly no difference between canonical and scrambled word order sentences in the implausible condition. On the other hand, in the plausible condition, L2 learners with lower proficiency (below zero) seemed to have more difficulty in answering scrambled sentences while L2 learners with higher proficiency scored very high on scrambled sentences (even better than canonical sentences). Thus, it was on plausible scrambled sentences that proficiency affected the L2 learners' accuracy.

3.3.6.2. Data trimming

Before analyzing reading time (RT) data, data trimming was performed and outliers were removed. In order to detect outlier observations, each person's overall mean and standard deviation in each reading region was calculated, and this was done for experimental items and filler items separately. Reading times 2.5 standard deviations above and below the mean RT were removed from further analysis. This led to the loss of 3.71% for NS data and 2.94% for L2 data of critical data points. There was no statistically significant effect on the number of outliers by condition (F= (3, 174) = 1.048, p = 0.373). At the same time, there was no statistically significant interaction between Condition and Group on the number of outliers (F = (3, 174) = 0.424, p = 0.736).

3.3.6.3. Reading time analysis

The critical regions were the second (second NP, scrambling effects) and third (embedded verb, plausibility and scrambling effects) and fourth (matrix subject, plausibility effects) regions. Since the length of words in the second region differed between sentences, residual reading times were computed for each group (NS and L2) by using a separate linear mixed effects model. Reading times from every region in all sentences, including the filler sentences from each group were included as input to each regression (Fine, Jaeger, Farmer, & Qian, 2013). The regression model had mean-centered word length as a fixed effect, and a random intercept for subject, and a by-subject random slope for mean-centered word length. The residuals of each model were saved as residual reading times for later analysis.

Analyses with all trials were conducted on residual reading times for each region (All answer analysis). At the same time, in order to see if there is any RT difference as a function of accuracy, only those trials that were answered correctly both in the online probe recognition task and offline agent identification task were included for a separate analysis (Correct answer analysis). This inclusion criterion for offline correct items was to make sure that L2 learners understood the meanings of the sentences, and the same inclusion criterion was used for native speakers. Thus, in total, 902 out of 960 (94%, Region 2), 899 out of 960 (94%, Region 3), 891 out of 960 (93%, Region 4) trials for native speakers and 1476 out of 1920 (77%, Region 2), 1479 out of 1920 (77%, Region 3), 1476 out of 1920 (77%, Region 4) trials for L2 speakers were analyzed for correct answer analyses. If the results of the two analyses (all answer analysis vs. correct answer analysis) differ from each other, the difference will be discussed. Otherwise, all answer analyses will be mainly discussed.

Linear mixed-effects models with random intercepts for subjects and items, and random slopes by subject and by item were performed separately on native speakers' and L2 learners' residual reading times at critical regions (second NP (Region 2), the embedded verb (Region 3), and the matrix subject (Region 4)) using *lme4* (Bates, Maechler, & Bolker, 2015) for modeling and *lmerTest* package for significance testing in R (version 3.4.3; R Core Team, 2017). As fixed effects, all models included word order and plausibility, and, for L2 models, cloze test scores and working memory capacity (WMC) scores.

Before modeling, categorical variables, word order and plausibility, were contrast-coded (-0.5 vs. 0.5 in alphabetical order). Two continuous variables for L2 models, cloze test scores and WMC scores, were mean-centered. All initial models for each region included maximal random effects structures with random intercepts, full random slopes (i.e., corresponding to the fixed effects and their interactions), and random correlations between slopes and intercepts (Barr, Levy, Scheepers, & Tily, 2013), but some models failed to converge. If models did not converge, a backwards model selection was performed to identify a final model that would converge. The resulting models fit to the residual RT data for each region included the maximal random effects structure supported by the data. The final random effects structure for each model and a full summary of the results are shown in Appendix B.

For native speakers, Figure 3.6 displays the observed reading times (raw RTs after data trimming) across all target sentences. The visual inspection of the graph reveals little RT difference between conditions in region 2 but clear RT differences in region 3. In region 3, native speakers took longest RTs in processing Implausible Scrambled (IC) sentences and shortest RTs in Plausible Canonical (PC) sentences and Plausible Scrambled (PS) and Implausible Canonical (IC) sentences between them. For region 4, the RT difference was small but implausible sentences seemed to have longer RTs than plausible sentences.

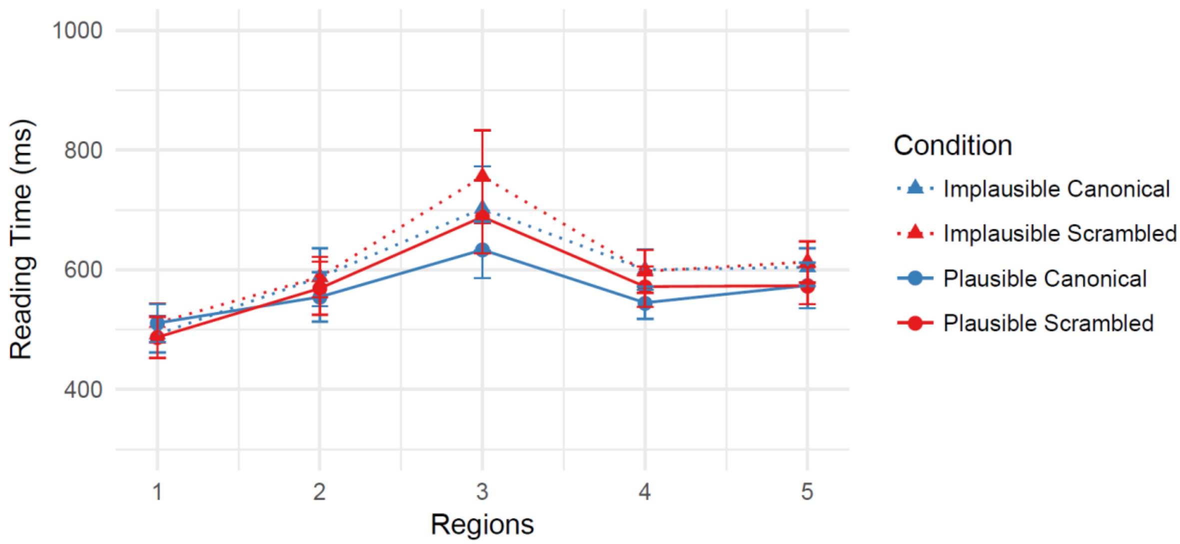

Figure 3.6. Native speakers' reading times in self-paced reading task

The statistical analysis shows that there was no significant effect of word order or plausibility at the second region (Plausibility: estimate = -26.535, SE= 14.685, $t(122.85) = -1.807$, $p= 0.073$; Word Order: estimate = 8.363, SE= 15.556, $t(51.39) = 0.538$, $p= 0.593$), which was visualized in Figure 3.6. At the third region (embedded verb), native speakers processed verbs more slowly in scrambled OSV sentences than in canonical SOV sentences (Word Order effect: estimate = 55.20, SE= 27.16, $t(34.40) = 2.032$, $p= 0.050$) and in implausible sentences than in plausible ones (Plausibility effect: estimate = -69.82, SE= 34.08, $t(33.50) = -2.049$, $p=$

0.048). However, there was no interaction between word order and plausibility (estimate = 1.94, SE= 46.05, $t(826.10) = 0.042$, $p= 0.966$), indicating that native speakers are sensitive to morphosyntactic information independently of plausibility. At the fourth region (matrix subject), native speakers continued to process implausible sentences more slowly than plausible sentences (Plausibility effect: estimate = -42.55, SE= 20.09, $t(28.45) = -2.118$, $p= 0.043$) without word order effects (estimate = 12.65, SE= 18.89, $t(26.22) = 0.670$, $p= 0.509$). No significant interaction between the two (estimate = 26.23, SE= 33.18, $t(56.11) = 0.791$, $p= 0.433$) was found. Analyses with correctly answered trials (correct answer analyses) for each region also had the same statistical results and will not be discussed further.

To sum up, for native speakers, scrambling effects were found at the embedded verb region, independently of plausibility of sentences. Even for native speakers, scrambled sentences were harder to parse than canonical word order sentences. At the same time, plausibility also influenced their reading times, in which implausible sentences had longer reading times than plausible ones. The native data also shows that the plausibility effects lasted longer than the word order effects. The former was still apparent in the next region (the matrix subject region) while the latter (scrambling effect) disappeared.

Results (raw RTs) from the 40 L2 learners are presented in Figure 3.7. Figure 3.7 shows different RT patterns from Figure 3.6 (NS data). The critical regions, the embedded verbs (region 3) and matrix subjects (region 4), did not seem to present any RT difference between conditions, while in region 2 (second NPs) Plausible Scrambled (OSV) sentences were read more slowly than other sentences. It should be noted that region 2 is where scrambling effects are expected but not plausibility, and plausibility effects are predicted at the embedded verb.

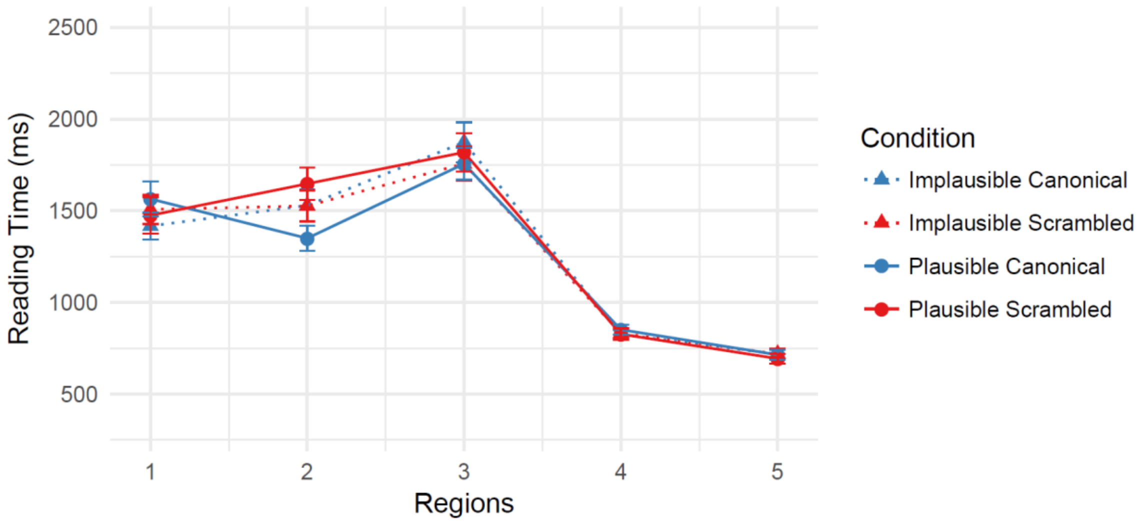

Figure 3.7. L2 speakers' reading times in self-paced reading task

The statistical analysis shows that there was a significant main effect of word order (estimate = 142.01, SE= 38.881, t(36.000) = 3.652, p= 0.001) at the second region. In addition, interactions were also observed, namely, a two-way interaction between word order and plausibility (estimate = 207.951, SE= 66.581, t(1758.200) = 3.123, p= 0.002), a two-way interaction between word order and WMC scores (estimate = -12.324, SE= 5.282, t(35.800) = -2.333, p= 0.025), and a three-way interaction among word order, cloze scores, and WMC scores (estimate =2.372, SE=1.005, t(36.400) =2.361, p=0.024). However, there were no significant main effects of plausibility (estimate = -28.987, SE= 38.695, t(35.200) = -0.749, p= 0.459), WMC scores (estimate = -6.137, SE= 5.406, t(36.000) = -1.135, p= 0.264), or cloze scores (estimate = -11.247, SE= 7.931, t(41.800) = -1.418, p= 0.164).

All the interactions found above were plotted into graphs using *emmeans* package, version 1.1.2 (Lenth, 2018) via *emmip* function to interpret the relationships.

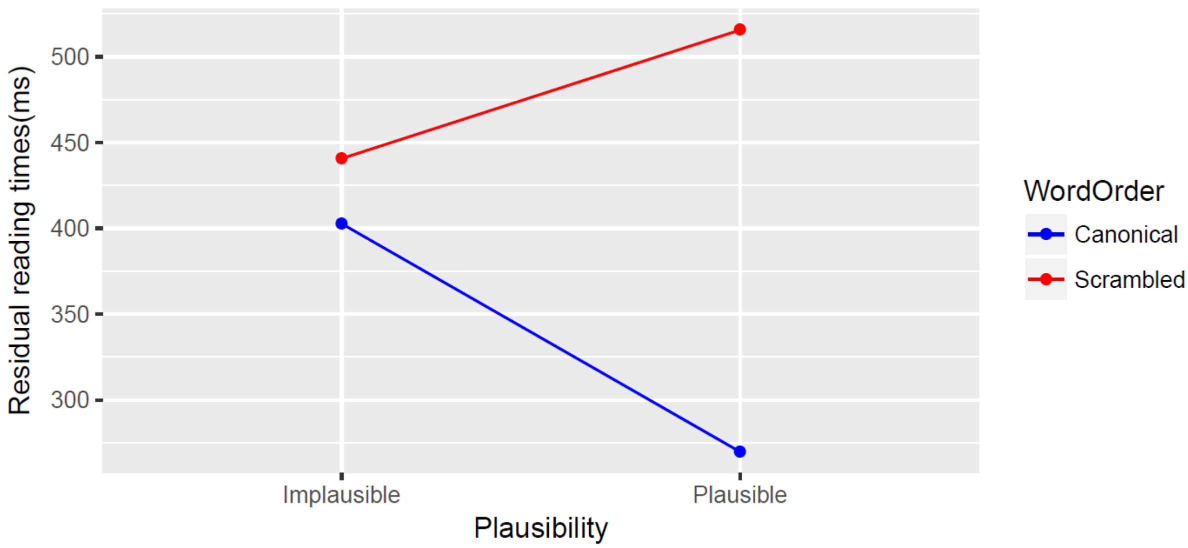

Figure 3.8. L2 interaction plot between word order and plausibility in region 2

The interaction graph between word order and plausibility shows that regardless of plausibility, scrambled sentences tended to have longer RTs. But, the difference in RTs between canonical and scrambled sentences was greater for plausible condition than for implausible condition.

The interaction between word order and WMC was also plotted into a graph as below. “WMC (centered)” on the x-axis indicates mean-centered WMC scores (0 equals the mean score).

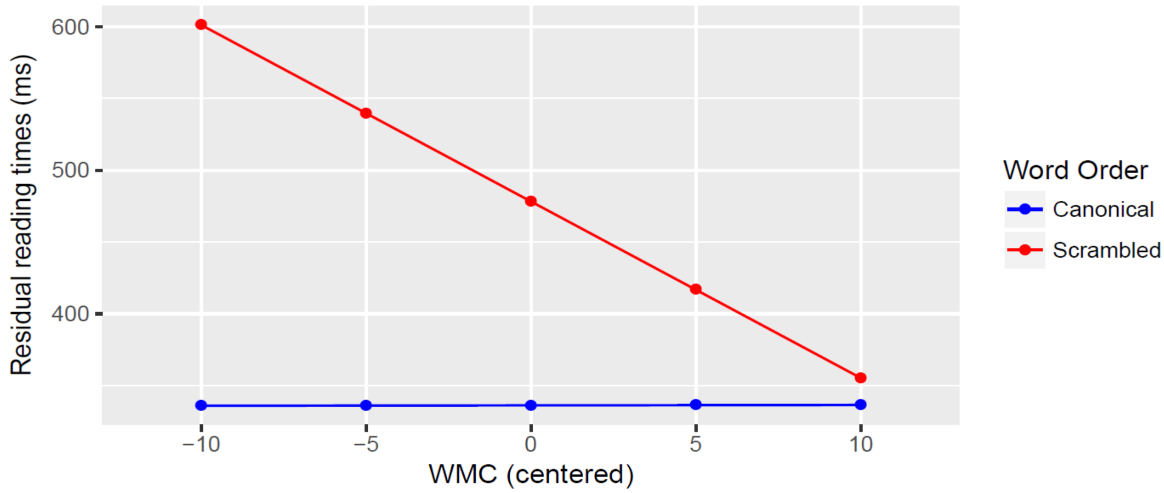

Figure 3.9. L2 interaction plot between word order and WMC in region 2

This interaction graph shows that WMC influenced processing of scrambled sentences while for canonical sentences, WMC did not affect L2 learners RTs. For L2 learners with greater WMC, the difference in RTs between canonical and scrambled sentences was small, while for L2 learners with smaller WMC, they tended to have much longer RTs in processing scrambled sentences than canonical word order sentences.

On the three-way interaction graph among word order, cloze scores, and WMC scores below, multiple panel plots were created on the combination of two continuous variables, mean-centered cloze scores and mean-centered WMC scores. The numbers on the top of each panel indicates the mean-centered cloze scores (0 equals the mean).

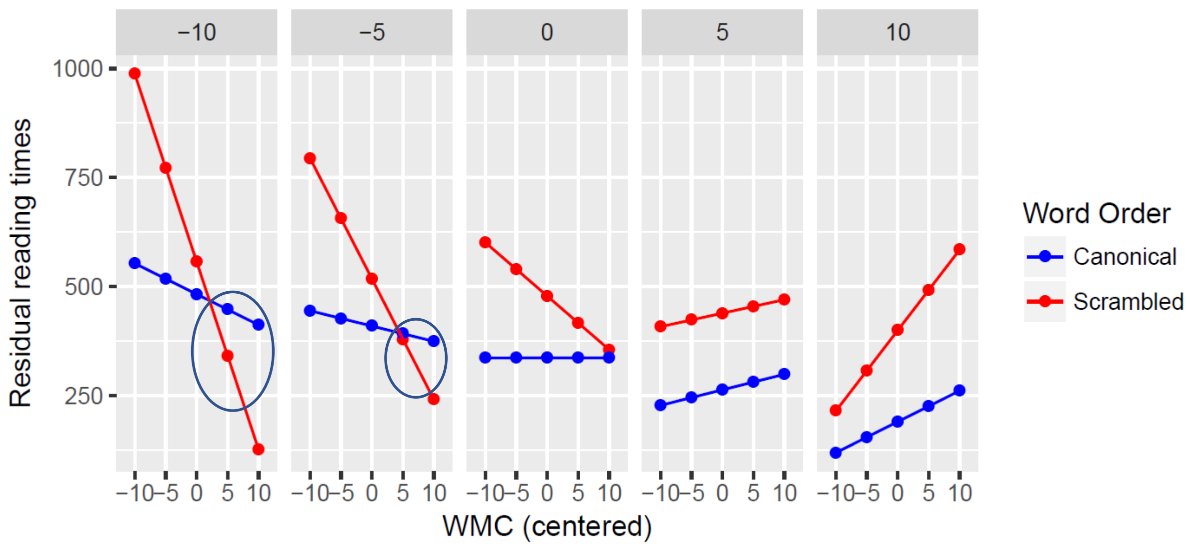

Figure 3.10. L2 interaction plot among word order, cloze scores, and WMC scores in region 2

The graph shows that most L2 learners behaved similarly to native speakers. That is, they tended to have longer RTs for scrambled sentences than canonical word order sentences. However, L2 learners with lower proficiency (below 0) and greater WMC (above 5) (as marked in circles on the graph) had longer RTs for canonical sentences than scrambled sentences. Thus, it is likely that advanced learners regardless of WMC exhibited longer reading times for the scrambled sentences than canonical word order sentences at this region, while less advanced learners showed less determinate reading time patterns.

For correct answer analysis in region 2, almost all the statistical results were the same as in all answer analysis. The only difference was the presence of two-way interaction between cloze scores and WMC scores (estimate = 2.078, SE= 0.995, $t(36.000) = 2.089$, $p = 0.044$) instead of the interaction between word order and WMC scores (estimate = -11.951, SE= 6.196, $t(43.700) = -1.929$, $p = 0.060$). Since the three-way interaction among word order, cloze scores, and WMC scores was discussed above, this two-way interaction will not be further discussed.

At the third region (embedded verb), no significant main effects were observed for plausibility (estimate =-26.265, SE=42.049, $t(37.500)$ =-0.625, p=0.536), word order (estimate =-22.956, SE=42.698, $t(35.600)$ =-0.538, p=0.594), cloze scores (estimate =3.701, SE=10.188, $t(36.100)$ =0.363, p=0.719), and WMC scores (estimate = -1.969, SE= 7.287, $t(36.100)$ = -0.270, p= 0.789). However, there were significant interactions, a two-way interaction between word order and plausibility (estimate = 180.266, SE= 76.138, $t(1612.700)$ = 2.368, p= 0.018) and a three-way interaction among word order, plausibility, and cloze scores (estimate =49.844, SE=14.881, $t(1686.000)$ =3.349, p=0.001).

For the two-way interaction between word order and plausibility, an interaction graph was plotted in Figure 3.11, using *emmeans* package via *emmip* function.

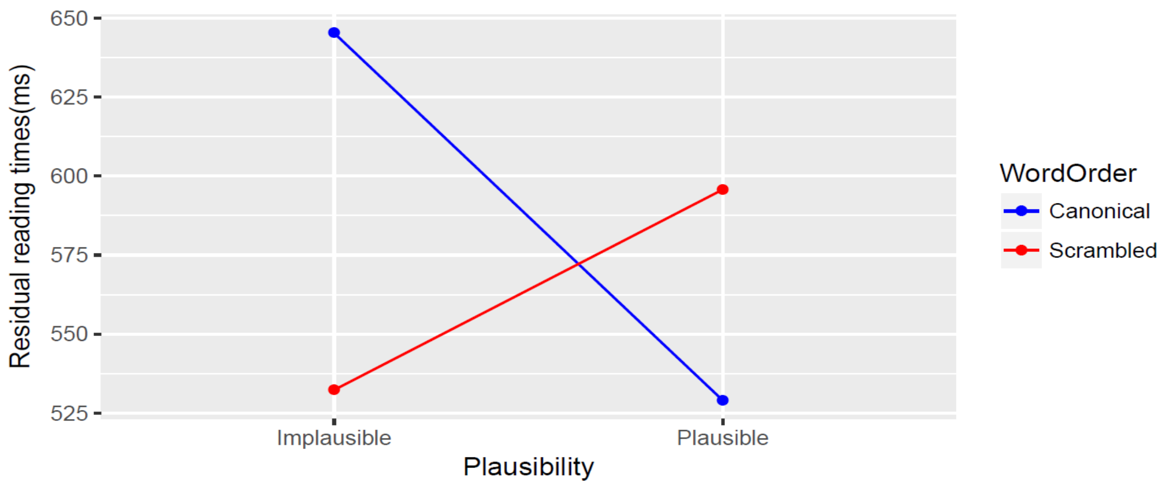

Figure 3.11. L2 interaction plot between word order and plausibility in region 3

The graph shows that L2 learners had longer RTs for scrambled sentences in plausible sentences, as predicted. However, in implausible sentences, the reverse RT patterns were observed. They had much longer RTs for canonical word order sentences than scrambled sentences, which is different from native RT patterns. In order to better understand the results, the significant three-way interaction among word order, plausibility, and mean-centered cloze scores (0 equals the average) was also plotted in Figure 3.12.

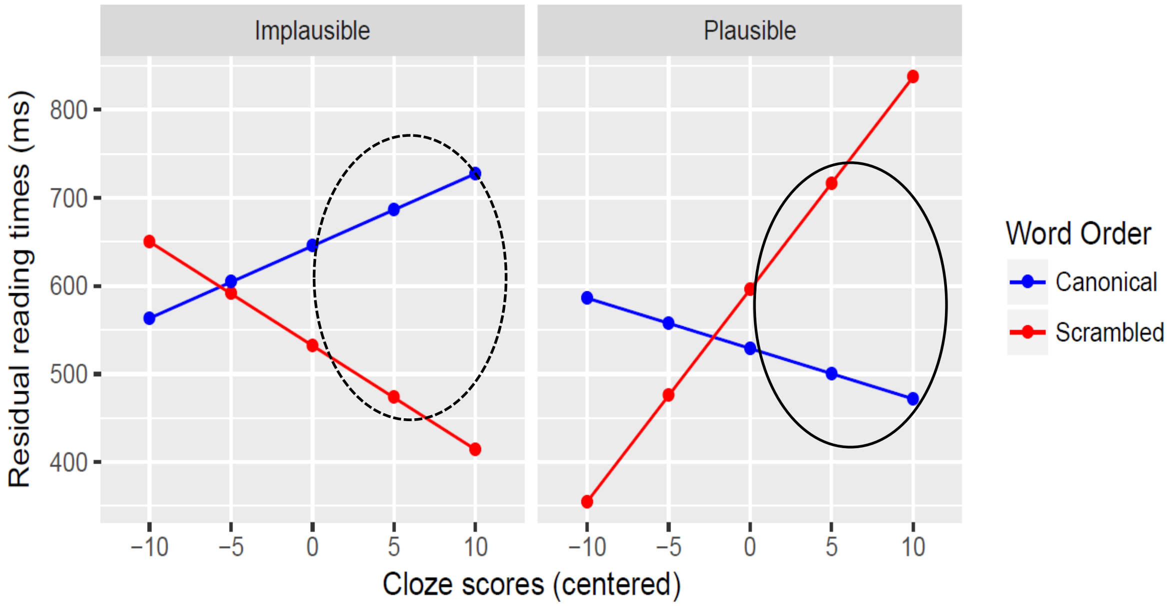

Figure 3.12. L2 interaction plot among word order, plausibility, and cloze scores in region 3

Looking at L2 learners with cloze scores better than the average (as marked in circles on the graph), they tended to have longer RTs for scrambled sentences than canonical word order sentences in plausible sentences (a circle in solid line). That is, if sentential meanings make sense according to their world knowledge, they tended to show similar RT patterns as native speakers. On the other hand, in implausible sentences, they tended to have longer RTs for canonical word order sentences than scrambled sentences (a circle in dotted line), which is the reverse pattern.

For L2 learners with cloze scores lower than the average, in plausible sentences, some tended to behave like higher proficiency learners but most of them tended to have longer RTs for canonical word order sentences. In implausible sentences, the majority tended to behave like higher proficiency learners and some had longer RTs for scrambled sentences. For lower proficiency learners below the average, they seem to have mixed RT patterns.

For this region (region 3), correct answer analysis also had the same statistical results as all answer analysis, so no further discussion will be given.

At the fourth region (matrix subject position), neither significant main effect nor interaction effect was observed in all answer analysis (Plausibility (estimate =8.960, SE=18.620, $t(33.600) = 0.481$, $p=0.633$); Word Order (estimate =-6.566, SE=16.093, $t(84.000) =-0.408$, $p=0.684$); Cloze scores (estimate =3.383, SE=5.863, $t(36.300) =0.577$, $p=0.567$); WMC (estimate =7.101, SE=4.211, $t(36.800) =1.687$, $p=0.100$) or correct answer analysis (Plausibility (estimate =11.997, SE=23.634, $t(33.100) =0.508$, $p=0.615$); Word Order (estimate = -3.991, SE=17.351, $t(232.200) =-0.230$, $p=0.818$); Cloze scores (estimate =3.538, SE=6.269, $t(35.700) =0.564$, $p=0.576$); WMC (estimate =8.371, SE=4.545, $t(37.700) =1.842$, $p= 0.073$)).

To summarize, for L2 learners, in region 2 (second NP), the majority of L2 learners tended to process scrambled sentences more slowly than canonical word order sentences. The effect of WMC was also observed in region 2 for scrambled sentences. The greater WMC the L2 learners had, the shorter RTs they took in this region. However, when the WMC effect was considered relative to proficiency, overall RT patterns were more complex, such that the majority still had longer RTs for scrambled sentences, while some showed the reverse RT patterns. In region 3 (embedded verb), proficiency played an important role in accounting for L2 RT patterns. First, L2 learners with higher proficiency (above the average) showed the scrambling effects in plausible sentences. However, even the higher proficiency learners showed the reverse patterns in implausible sentences. L2 learners with lower proficiency (below the average) showed mixed RT patterns in this region. Some behaved similarly to higher proficiency L2 learners, but others did not. Considering proficiency effects, the high proficiency learners showed clear scrambling effects in region 2, and they continued to show the same RT patterns in plausible sentences but not in implausible sentences in region 3. On the other hand, the low proficiency learners showed mixed RT patterns in regions 2 and 3.

3.4. Discussion

In Experiment 1, processing of (plausible or implausible) sentences with local scrambling was examined with native Korean speakers and L1 Mandarin Chinese-L2 Korean speakers. In this study, L2 learners' proficiency and working memory capacities were considered to investigate individual differences in L2 sentence processing. Results from online and offline tasks revealed different response patterns between native and L2 speakers.

For the offline agent identification task, results from native speakers did not show any significant effect of word order or plausibility. Native speakers scored almost at ceiling, and these two factors did not affect their accuracy. In contrast, results from L2 learners showed plausibility effects as well as proficiency effects: implausible sentences were harder than plausible ones, and L2 learners with higher proficiency tended to score better on the offline task. Regarding word order, if implausible sentences were scrambled, the L2 learners had more difficulty interpreting the sentences. However, if sentential meanings were supported by their world knowledge, they succeeded in interpreting the scrambled sentences even more accurately than the canonical ones. In other words, plausibility had a facilitative effect on the offline processing of scrambled sentences.

For the online self-paced reading task, the critical regions were region 2 (second NP), region 3 (embedded verb), and region 4 (matrix subject). Native Korean speakers did not show any effect of word order or plausibility in region 2. In region 3, however, they processed verbs more slowly in OSV (scrambled) sentences than in SOV (canonical) sentences (scrambling effects). They also showed plausibility effects (longer RTs for implausible than plausible sentences) regardless of word order. The results from native speakers indicate that local scrambling involves an additional processing load at the verb, which can be interpreted as due to the effect of processing of the gap (RQ1), predicted by the movement account of local scrambling. In addition, providing plausibility information does not eliminate scrambling effects, although it reduces the magnitude of the effect, which means native speakers used not only case-marking information but also plausibility information in processing scrambled sentences (RQ2b). Since their RT patterns in SOV sentences were parallel to OSV sentences with respect to plausibility effects, we can also assume that the native speakers possibly used both case-marking and plausibility information when they processed SOV sentences (RQ2a).

For L2 learners, proficiency effects but not WMC effects were apparent in their online sentence processing (RQ4). L2 learners with higher proficiency, established by a cloze test, showed scrambling effects in region 2, similar to native RT data in region 3 (RQ1). In region 3, higher proficiency L2 learners showed consistent scrambling effects in plausible sentences but not in implausible ones, different from native speakers. L2 learners with lower proficiency showed mixed RT patterns in both regions 2 and 3 (less determinate RT patterns), that is, some behaved similarly to higher proficiency L2 learners but other did not.

Based on the online results from the L2 learners, even though L2 learners with higher proficiency showed scrambling effects (longer RTs for OSV than SOV sentences) in region 2 like native speakers, when the plausibility of sentences became obvious in region 3, their RT patterns indicate their reliance on heuristic word order information and plausibility (Prediction 2). According to Prediction 2, for OSV scrambled sentences, if L2 learners use the heuristic word order and plausibility cues, and disregard case marking, they would misparse OSV (scrambled) sentences as SOV (canonical) ones. This predicts that OSV scrambled plausible sentences like (3) would be read more slowly than OSV scrambled implausible sentences like (4) (the target sentence examples are repeated below).

(1) Condition A: canonical, plausible

sensayngnim-i	haksayng-ul	kaluchyessta-ko	John-i	sayngkakhayssta
teacher-NOM	student-ACC	taught-COMP	John-NOM	thought

'John thought that the teacher taught the student.'

(2) Condition B: canonical, implausible

haksayng-i	sensayngnim-ul	kaluchyessta-ko	John-i	sayngkakhayssta
student-NOM	teacher-ACC	taught-COMP	John-NOM	thought

'John thought that the student taught the teacher.'

(3) Condition C: scrambled, plausible

haksayng-ul	sensayngnim-i	kaluchyessta-ko	John-i	sayngkakhayssta
student-ACC	teacher-NOM	taught-COMP	John-NOM	thought

'John thought that the teacher taught the student.'

(4) Condition D: scrambled, implausible

sensayngnim-ul	haksayng-i	kaluchyessta-ko	John-i	sayngkakhayssta
teacher-ACC	student-NOM	taught-COMP	John-NOM	thought

'John thought that the student taught the teacher.'

Looking at reading time patterns, this prediction is confirmed by higher proficiency L2 speaker data (RQ2b). In addition, their RT patterns in SOV sentences were consistent with our prediction as well. Implausible sentences had longer RTs than plausible ones (plausibility effects). Since it is possible for the L2 learners to parse SOV sentences accurately even if they rely on heuristic word order alone, the result does not provide strong evidence to answer our research question RQ2a. However, at least it shows that the L2 learners had sufficient proficiency to be able to process canonical word order sentences accurately like native speakers.

Overall, depending on proficiency levels, less proficient L2 learners did not show determinate reading time patterns, but even higher proficient L2 learners performed differently from native speakers (RQ2c). The results could be interpreted as learners' inability to use case-marking information in the online sentence processing. However, it should be also noted that their offline results from the agent identification task showed their use of case-marking information in sentence interpretations. Then, it would be also possible to consider the different response patterns due to the differences between online and offline tasks.

Before drawing a conclusion about differences between native and L2 processing as well as online and offline L2 processing, however, we first need to discuss a limitation of our online self-paced reading task. Since comprehension questions were not given (instead, probe recognition questions were presented), it is still unclear whether the L2 learners could get the complete, correct final interpretation for each sentence in online sentence processing. Since the task did not necessarily require them to get the complete interpretations to answer the questions, they might not have paid enough attention to the case-marking information when plausibility conflicted with case-marking cues. Since scrambled sentences were harder to parse even for native speakers, L2 learners could have settled on incomplete interpretations because they did not have to get the complete interpretations, even though they had an ability to do so.

Thus, it is still inconclusive whether L2 learners' reading time patterns are due to their inability to use case-marking information in online sentence processing or task effects. It is necessary to control the task effects to draw more conclusive answers to the research questions. Motivated by such unanswered questions in Experiment 1, Experiment 2 was designed to address the same research questions further, investigating long-distance scrambling and targeting more proficient L2 learners, by considering the task effects.

Chapter 4

Experiment 2: Processing of Long-Distance (LD) scrambling

As in Experiment 1, the word order and plausibility of sentences were manipulated in Experiment 2 to examine how scrambling influences L1 and L2 sentence processing. Different from Experiment 1 (local scrambling), Experiment 2 tests long-distance (LD) scrambling, where the distance between the moved element (filler) and the posited gap in scrambled sentences was manipulated. Using a self-paced reading task, Experiment 2 was carried out to answer the following research questions.

RQ1. Does (local and long-distance) scrambling increase the processing load in L1 and/or L2 sentence processing?

RQ2. How do native and L2 speakers process canonical and scrambled sentences using real world plausibility as a heuristic?

2a. Do native and L2 speakers use case marking information and/or non-syntactic cues in canonical sentences?

2b. Do native and L2 speakers use case marking information and/or non-syntactic cues in scrambled sentences?

2c. Are native and L2 speakers the same or different with regard to 2a and 2b?

RQ4. Do L2 learners' proficiency and/or WMC play a role in their sentence processing? If so, is there any relationship between proficiency and working memory in L2 sentence processing?

4.1. Methodology

4.1.1. Procedure (Task order)

Experiment 2 was conducted in a similar way to Experiment 1. The main task was a self-paced reading task, in which a comprehension question followed each sentence unlike the self-paced reading task combined with a probe recognition task in Experiment 1. After the online reading task, an offline task, an agent identification task, was administered as in Experiment 1 to test L2 learners' offline knowledge of word order and case markers in Korean. On the other hand, a sentence recognition task was not used since the comprehension questions were aimed to the same purpose, making participants pay attention to the meaning of the sentences. After the offline task, a cloze test and a language background questionnaire that were used in Experiment 1 were given to both native and L2 speakers. Lastly, a shortened version of three working memory capacity tasks, used only for L2 speakers in Experiment 1, were administered to both native and L2 speakers. After completing all the tasks, only L2 speakers were given a vocabulary test to examine their lexical knowledge of target sentences in the self-paced reading task.

For NS & L2 [1. Self-paced reading task] ⇨ [2. Agent identification task]
⇨ [3. Cloze task] ⇨ [4. Language background questionnaire]
⇨ [5. Working memory capacity task] Only L2 ⇨ [6. Vocabulary task]

Figure 4.1. The order of tasks for Experiment 2

4.1.2. Participants

24 native Korean speakers who did not participate in Experiment 1 were included as a baseline group, and they were all tested in Korea. 40 L1-Mandarin Chinese L2-Korean speakers

participated in this study as a target group, and they were also recruited and tested in Korea. In Experiment 1, the target population was L2 learners at intermediate to advanced proficiency levels whereas in Experiment 2, advanced (or highly advanced) L2 learners were included since lower proficiency L2 learners in Experiment 1 showed less determinate reading time patterns. Thus, the inclusion criteria for L2 speakers was slightly modified from the ones from Experiment 1 to include more advanced learners. In Experiment 2, the learners had to study Korean at least for four years or live in Korea as a college student at least for four years or have a TOPIK certificate of level 5 (advanced level) or higher. Except for learners' proficiency levels, the rest of the criteria remained the same as in Experiment 1. Learners' proficiency levels were additionally measured by using the same cloze test as in Experiment 1. Table 4.1 below summarizes the language background information for both native and L2 speakers in Experiment 2.

Table 4.1. Background information for the participants in Experiment 2

Group	Total number	Age	First age of instruction	Length of study Korean (Months)	Length of residence in Korea (Months)	Cloze scores (Total: 40)
NS	24 (Male: 7 Female: 17	Mean: 23 Range: 20~29	N/A	N/A	N/A	Mean: 38.5 SD: 1.4 Range: 33~40
L2	40 (Male: 5 Female: 35)	Mean: 25 Range: 21~28	Mean: 20 Range: 19~24	Mean: 33.9 Range: 12~54	Mean: 27.3[11] Range: 1 week~72 mons	Mean: 29.5 SD: 3.8 Range: 22~36

[11] If the learners have lived in Korea less than a month, their length of residence was calculated by dividing the number of weeks by four (weeks).

Compared to the mean score (26.6) on the cloze test for the L2 learners in Experiment 1, the L2 learners in Experiment 2 had a higher average score with a smaller standard deviation on the same cloze test, indicating they were more advanced learners than the learners in Experiment 1.

4.1.3. Materials

In Experiment 2, the distance between fillers and gaps were manipulated to see whether L2 learners of Korean can process scrambled sentences with longer distance filler-gap dependencies, compared to Experiment 1.

As in Experiment 1, 48 sets of four bi-clausal sentences were used for the self-paced reading task in Experiment 2 by manipulating word order (canonical vs. scrambled) and plausibility (plausible vs. implausible) like (1) to (4). First, LD scrambling sentences were created by moving the embedded object out of the embedded clause to the sentence-initial position as in (3) and (4). Plausibility was also manipulated as in Experiment 1 by switching the word order between the embedded subject and object of the embedded verb like (1) and (2). The experimental items were counterbalanced across four lists, so that each participant read only one token from each set.

(1) Condition A: canonical, plausible

Region1	**Region 2**	**Region 3**	**Region 4**	**Region 5**
na-nun	ecey	sensayngnim-i	kyosil-eyse	haksayng-ul
I-TOP	yesterday	teacher-NOM	in the classroom	student-ACC
Region 6	**Region 7**	**Region 8**	**Region 9**	
yelsimhi	kaluchyessta-ko	ttodasi	sayngkakhayssta	
enthusiastically	taught-COMP	again	thought	

‘I thought again that the teacher taught the student enthusiastically in the classroom yesterday.’

(2) Condition B: canonical, implausible

na-nun	ecey	haksayng-i	kyosil-eyse	sensayngnim-ul
I-TOP	yesterday	student-NOM	in the classroom	teacher-ACC
yelsimhi	kaluchyessta-ko	ttodasi	sayngkakhayssta	
enthusiastically	taught-COMP	again	thought	

‘I thought again that the student taught the teacher enthusiastically in the classroom yesterday.’

(3) Condition C: scrambled, plausible

haksayng$_i$-ul	na-nun	ecey	sensayngnim-i t_i	kyosil-eyse
student-ACC	I-TOP	yesterday	teacher-NOM	in the classroom
yelsimhi	kaluchyessta-ko	ttodasi	sayngkakhayssta	
enthusiastically	taught-COMP	again	thought	

‘I thought again that the teacher taught the student enthusiastically in the classroom yesterday.’

(4) Condition D: scrambled, implausible

sensayngnim$_i$-ul	na-nun	ecey	haksayng-i t_i	kyosil-eyse
teacher-ACC	I-TOP	yesterday	student-NOM	in the classroom
yelsimhi	kaluchyessta-ko	ttodasi	sayngkakhayssta	
enthusiastically	taught-COMP	again	thought	

‘I thought again that the student taught the teacher enthusiastically in the classroom yesterday.’

The embedded clause of target sentences was the same as in Experiment 1[12]. To generate LD scrambled sentences, the matrix subject appeared in the sentence initial position in canonical sentences, so that the object of the embedded verb could be scrambled over a clause. In addition,

[12] The embedded subject and object in one of 48 sets from Experiment 1 were modified in Experiment 2 for a better interpretation.

each sentence had four adverbial modifiers after the matrix subject, the embedded subject, the embedded object, and the embedded verb, respectively, to allow for spillover effects. Since the modifiers could influence the plausibility, a separate norming task was administered to confirm the plausibility of the sentences and the results will be discussed in session 4.3.1.

Due to the manipulation of sentences, the target sentences in Experiment 2 were inevitably temporarily ambiguous by having the matrix subject and the embedded subject next to each other, which was also different from Experiment 1. In order to facilitate reanalysis of garden-path sentences, the matrix subject was always *na* 'I' and marked by a topic marker *–nun* since the topic marker *–nun* is likely to be attached to the matrix subject (Kim, 1999).

For the filler sentences, 120 filler items with different syntactic structures such as subject/object relative clauses, coordinate constructions, adverb-initial sentences, and sentences with other types of non-nominative initial NP were included in order to discourage the development of a particular strategy (for example, if you see an accusative marked NP, you just need to wait for the nominative-marked NP). The plausibility of filler sentences was manipulated as in Experiment 1. The filler sentences were presented with six, seven, or eight regions, to inhibit a particular parsing strategy (such as looking for the 7th region as the disambiguating region). In total, 12 practice items and 168 (48 target and 120 filler) sentences were presented (see Appendix A).

4.1.4. Self-paced reading task

Same as in Experiment 1, the main task in Experiment 2 was a self-paced reading task. Unlike Experiment 1, however, comprehension was tracked via a comprehension question after each sentence rather than a probe recognition task. The comprehension questions asked about either subject or object of the embedded verb for target sentences, while for filler sentences, mixed questions such as asking about an agent, a theme, a verb, or an adverb were given to the participants.

For example, participants read a sentence like (1), one region at a time, and a comprehension question like (5), either Yes or No Type question, was given to them. They had to answer the question by pressing YES or NO button. Half of the questions had target answers of YES, and the other half had target answers of NO.

(5) **Yes Type Question**

선생님이 학생을 가르쳤습니까? 예 아니요

"Did the teacher teach the student?" Yes No

No Type Question

학생이 선생님을 가르쳤습니까? 예 아니요

"Did the student teach the student?" Yes No

Although comprehension questions for target sentences were asking explicitly about what was being tested, the comprehension questions were necessary to check participants' final interpretation of sentences to ameliorate the possible task effects discussed in Experiment 1, and filler questions were expected to help disguise the main purpose of the study. The rest of the procedure was the same as in Experiment 1.

4.1.5. Offline Tasks

One offline task, an agent identification task, was administered to participants. The same sentences and comprehension questions (without fillers) as in the online task were given to examine whether L2 learners had knowledge of word order and case markers in Korean. Since the offline task was not time-constrained, the task was also used to examine whether the learners would perform better in the offline task with more time available. In total, four practice items and 48 target sentences were presented.

After the completion of the offline task, both native and L2 speakers were given the same cloze task and language background questionnaire as in Experiment 1.

4.1.6. Working Memory Capacity Tasks

The same task was given as in Experiment 1 to both native and L2 speakers (see 3.1.6).

4.1.7. Vocabulary Task

Only L2 learners took a multiple-choice vocabulary test to ensure their lexical knowledge to complete the online task. 68 nouns and 42 verbs used in the online task were tested. The rest of the procedure was identical to that in Experiment 1.

4.2. Predictions

The critical regions for observing word order and plausibility effects were region 4 (scrambling effects), region 5 (spill-over region), region 7 (plausibility effects), and region 8 (spill-over region).

Before discussing the critical regions, other possible effects before the critical regions are predicted. In canonical word order sentences as in (1) and (2), native speakers would possibly slow down at the second NP (NP-NOM) when they realize that the second NP (region 3) cannot be an argument of the same verb as the first NP (I-TOP) in the same clause. In scrambled sentences as in (3) and (4), native speakers would probably slow down at the second NP (region

2). At the second NP, they are expected to slow down due to trying to incorporate the fronted accusative marked NP with the topic marked *I*. However, these effects are not the focus of the study and not controlled for comparison between conditions. Thus, they will not be further discussed.

At the third NP (region 4) in scrambled sentences, when native speakers encounter the nominative marked NP, they are expected to slow down due to realizing that the initial analysis (incorporating the accusative marked NP with the topic marked *I*) needs to be revised (garden path effects) and parsing the scrambled element as a part of the embedded clause (scrambled effects). In other words, they need to reanalyze the accusative marked NP as a co-argument of the nominative marked NP instead of the topic marked *I*. Since the scrambling effects and/or the garden path effects are expected in region 5 (spill-over region) as well, regions 4 and 5 in scrambled sentences are expected to have slower reading times than in canonical word order sentences, if scrambling affects their reading. As for plausibility effects, native speakers are expected to process embedded verbs (region 7) more slowly in implausible sentences than in plausible ones, regardless of word order, as observed in Experiment 1.

For L2 speakers, since there are three NPs in a row before the embedded verb, it is expected that it will be difficult for them to use a heuristic (canonical) word order strategy (assuming the first NP as the subject and the second as the object) even in canonical word order sentences like (1) and (2). Thus, it is predicted that if L2 speakers can use case marking information, they will show plausibility effects at the embedded verb in canonical word order sentences. Unlike in Experiment 1, a (canonical) word order strategy cannot provide learners with the correct parse in Experiment 2. Learners have to use case-marking information to show sensitivity to plausibility.

For scrambled sentences like (3) and (4), if L2 learners can rely on case-marking information, they are expected to be able to show the same garden path and scrambling effects in regions 4 and 5 like native speakers. However, if they do not, or get confused with the three NPs, or delay their interpretations until a verb is given, they are expected not to show any differences in reading times between scrambled and canonical sentences in regions 4 and 5. At region 7 (embedded verb), if they use case-marking information, they will show sensitivity to plausibility with longer RTs for implausible sentences than plausible sentences. If they do not, and fail to incorporate the topic-marked *I* into the structure, and rely on the relation between the subject and the object of the embedded verb (i.e., world knowledge or plausibility), a scrambled implausible sentence like (4) should be read similarly to a canonical plausible sentence like (1), and a scrambled plausible sentence like (3) to a canonical implausible sentence like (2) at the embedded verb. In other words, if learners fail to use case-marking information but parse the sentences based on plausibility, they are expected to read scrambled plausible sentences more slowly than scrambled implausible sentences. Table 4.2. summarizes the predictions as below.

Table 4.2. Predictions of reading time patterns at region 7 (embedded verb)

Predictions (qualitative similarity/difference)	Reading time patterns (A, B, C, and D refer to target sentence conditions)
Prediction 1. L1 processing = L2 processing	NS: $A < B \leq C < D$
	L2: $A < B \leq C < D$
Prediction 2. L1 processing ≠ L2 processing	NS: $A < B \leq C < D$
	L2: $A \leq D < C \leq B$

A: (SOV) canonical, plausible; B: (SOV) canonical, implausible;
C: (OSV) scrambled, plausible; D: (OSV) scrambled, implausible

4.3. Results

4.3.1. Plausibility norming task

Procedures (data collection and scoring) were identical to those used in Experiment 1. Four native Korean speakers who did not participate in the main study completed the norming task, in which 48 experimental and 98 filler sentences were given.

A paired samples t-test was performed on the averaged z-scores for plausible and implausible sentence pairs. Significant differences were observed between the plausible and implausible sentences in each set ($t(47) = 33.793$, $p<.001$) as well as across the three types of plausibility relations (human: $t(15) = 24.441$, $p<.001$; human/animal: $t(15) = 16.259$, $p<.001$; animal: $t(15) =22.462$, $p<.001$). Thus, all the experimental sentences were retained for the main study.

4.3.2. Agent identification task

When participants were asked to answer questions about either subject or object of the embedded verb, native Korean speakers performed almost at ceiling, (99% correct (1140 out of 1152)). The accuracy rates between conditions is presented in Table 4.3.

Table 4.3. Descriptive statistics results of agent identification task for Experiment 2

Group	Word Order	Plausibility			
		Plausible		Implausible	
		Proportion Correct	SD	Proportion Correct	SD
NS	Canonical	0.993	0.024	0.997	0.017
	Scrambled	0.983	0.049	0.986	0.032
L2	Canonical	0.948	0.148	0.879	0.162
	Scrambled	0.915	0.143	0.760	0.294

For L2 learners, the overall accuracy rate across the four conditions was 88% (1681 out of 1920). Although the accuracy rate for the implausible scrambled (IS) sentences was lower than other conditions, at least the majority of the L2 learners (25 out of 40) were able to achieve accuracy higher than 80% in this condition. In order to examine what affects such difference in accuracy between conditions, statistical analyses were conducted using the identical procedure to that in Experiment 1 (see Appendix B for the final random effects structure for each model and a full summary of the results).

For native speakers, there were no significant main effects of plausibility (estimate = 0.188, SE=1.142, z=0.165, p=0.869) or word order (estimate =-1.172, SE=0.703, z=-1.667, p=0.096), nor any significant interaction between them (estimate =0.455, SE=1.406, z=0.323, p=0.747). These results suggest that the two factors did not affect the accuracy rates of native speakers on the offline agent identification task.

For L2 learners, significant main effects of plausibility (estimate = 2.555, SE= 0.924, z = 2.764, p= 0.006), word order (estimate = -2.326, SE= 0.915, z = -2.543, p= 0.011), and cloze scores (estimate = 0.376, SE= 0.122, z = 3.094, p= 0.002) were observed. In addition, a significant two-way interaction between plausibility and word order (estimate =-3.617, SE=1.750, z =-2.067, p= 0.039) and a significant three-way interaction among plausibility, word order, and cloze scores (estimate =-0.626, SE=0.245, z =-2.551, p=0.011) were also detected in this analysis.

These results show that the accuracy rates for the L2 learners were higher in plausible than in implausible sentences and in canonical than in scrambled sentences, as shown in Table 4.3. Besides that, proficiency also affected the L2 accuracy. In order to examine the relationship among the three factors, plausibility, word order, and mean-centered cloze scores (0 equals the mean), the three-way interaction among them was plotted using *effects* and *ggplot2* packages as below.

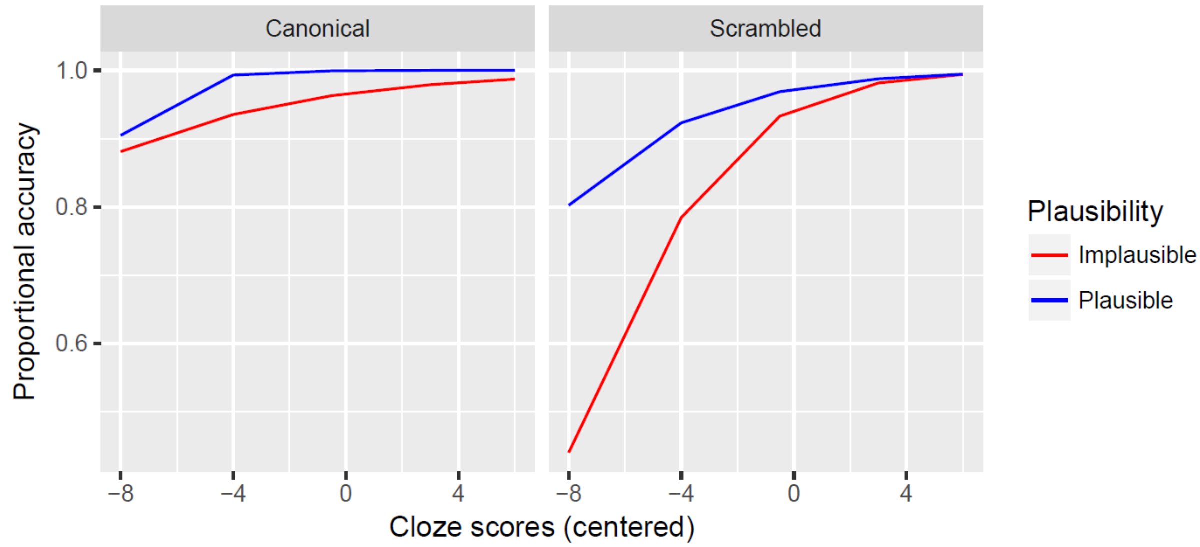

Figure 4.2. L2 interaction plot among plausibility, word order, and cloze scores on the agent identification task

The graph on the left shows that implausible sentences in canonical word order were harder than plausible ones across all proficiency levels. However, the difference in L2 accuracy between the two levels of plausibility was relatively small, compared to the one found in scrambled sentences. In the scrambled sentences, the graph on the right shows that not only plausibility but also learners' proficiency influenced the L2 accuracy. Depending on the proficiency levels, less proficient learners (below zero on the x-axis) were much less accurate in answering questions for implausible sentences than for plausible sentences, while more proficient L2 leaners (above zero on the x-axis) were able to interpret the implausible sentences as accurately as plausible ones. Overall, although the implausible scrambled sentences were generally hard for L2 learners, native-like performances in the offline comprehension task can be achieved by learners with sufficient L2 proficiency.

4.3.3. Working Memory Capacities

Data collection and scoring were identical to that in Experiment 1 except for that both native and L2 speakers completed the WMC tasks (Operation Span task (Ospan), Symmetry Span task (Sspan), and Rotation Span task (Rspan)) in Experiment 2. From each task, two scores, accuracy rates on distractor items and (partial) span scores on recall items, were obtained.

For native speakers, the average distractor accuracy was 22.5 out of 25 (SD=1.8) for Ospan (90% correct), and 13.1 out of 14 (SD=1.1) for Sspan (93% correct), and 12.1 out of 14 (SD=1.4) for Rspan (87% correct). For L2 speakers, the corresponding values were 21.2 out of 25 (85% correct) (SD=2.5), 12.4 out of 14 (89% correct) (SD=2.5), and 12.1 out of 14 (87% correct) (SD=1.8), respectively.

Descriptive statistics for the partial span scores by native and L2 speakers are displayed in Table 4.4. It should be noted that in all further analyses, WMC scores were the sum of the partial scores from the three tasks, as in Experiment 1.

Table 4.4. WMC scores of three WMC tasks

Group	Ospan (total: 25)	Sspan (total: 14)	Rspan (total: 14)	WMC (total: 53)
NS	Average: 20.3 SD: 4.6 Range: 9~25	Average: 11.2 SD: 2.3 Range: 6~14	Average: 8.9 SD: 2.1 Range: 4~13	Average: 40.4 SD: 4.7 Range: 29~46
L2	Average: 13.8 SD: 4.5 Range: 3~23	Average: 9.2 SD: 2.4 Range: 5~14	Average: 8.8 SD: 2.5 Range: 4~14	Average: 31.7 SD: 6.3 Range:14~45

4.3.4. Vocabulary task

Same as in Experiment 1, L2 learners' lexical knowledge was measured by using a vocabulary task. The average score of 40 L2 speakers on the task was 104.1 out of 110 (95% correct) with a standard deviation of 2.93, which indicates that they scored better in comparison to the L2 learners in Experiment 1 (91% correct). Thus, the same inclusion criterion as in Experiment 1 was applied to Experiment 2 as well, by having all items for further analyses.

4.3.5. Self-paced reading task

4.3.5.1. Online comprehension task

Before discussing reading time data, online comprehension results need to be discussed to compare the accuracy on the online task to that on the offline task. Table 4.5. summarizes the results of both native and L2 speakers.

Table 4.5. Online comprehension accuracy

Group	Word Order	Plausibility			
		Plausible		Implausible	
		Proportion Correct	SD	Proportion Correct	SD
NS	Canonical	0.944	0.059	0.889	0.151
	Scrambled	0.938	0.071	0.753	0.220
L2	Canonical	0.827	0.126	0.462	0.192
	Scrambled	0.758	0.179	0.362	0.259

The descriptive statistics show that even native speakers had comparatively lower accuracy in implausible sentences than in plausible ones. For L2 learners, implausible sentences were very difficult for them to process, as shown by less than 50% accuracy in the implausible condition. Although the learners were able to comprehend scrambled sentences with plausible meanings (76% correct), they had difficulty in processing canonical sentences with implausible meanings (46% correct).

Statistical analyses were conducted including working memory capacity as a factor to examine what affected online comprehension accuracy in both native and L2 speakers. Since fixed and random effects, as well as a dependent variable (accuracy data (i.e., binary), were the same as in the probe recognition task in Experiment 1, the modeling procedures were identical to those used in the probe recognition task except for that mean-centered WMC scores were included as a fixed effect (see Appendix B for the final random effects structure for each model and a full summary of the results).

For native speakers, there was a significant main effect of WMC (estimate=-0.128, SE=0.054, *z*=-2.399, *p*=0.016). Individual differences in WMC were negatively associated with native speakers' online accuracy, such that native speakers with higher WMC had lower accuracy than those with lower WMC on the online comprehension task. In addition to the effect of WMC, a significant main effect of plausibility (estimate=1.476, SE=0.524, *z*=2.818, *p*=0.005) as well as a significant two-way interaction between plausibility and word order (estimate=1.944, SE=0.943, *z*=2.062, *p*=0.039) were observed. The effect of plausibility was interpreted with consideration of the interaction, and its relationship with word order was plotted by using *emmeans* package via *emmip* function.

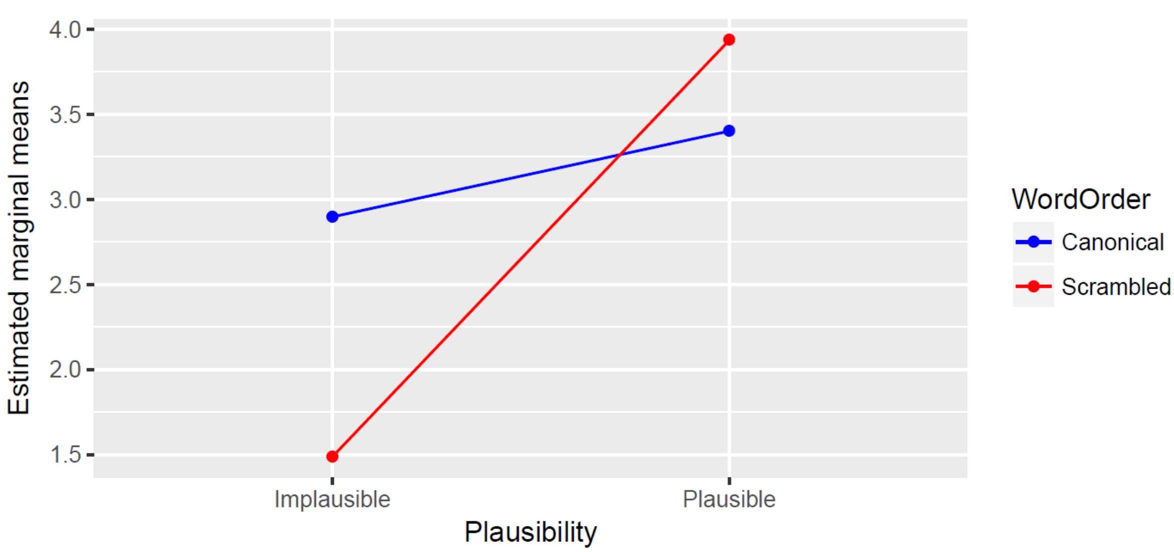

Figure 4.3. NS interaction plot between plausibility and word order on the online comprehension task. Note that the values on the y-axis are estimated marginal means of accuracy on the logit-scale from the fitted L2 model.

Interestingly, native speakers were more accurate in scrambled sentences than canonical sentences in the plausible condition, while the reverse accuracy patterns were observed in the implausible condition. Thus, the results from native speakers show that plausibility effects were more apparent in scrambled sentences than in canonical sentences.

For L2 speakers, significant main effects of plausibility (estimate=1.988, SE=0.180, z=11.051, p< .001), word order (estimate=-0.412, SE=0.145, z=-2.840, p=0.005), and cloze scores (estimate=0.108, SE=0.031, z=3.444, p=0.001) were obtained without any significant interaction effects between them. The results show that L2 learners responded more accurately to plausible sentences than to implausible sentences, and to canonical sentences than to scrambled sentences. At the same time, L2 proficiency influenced accuracy: more advanced L2 learners comprehended sentences more accurately than less advanced L2 learners.

To sum up, during online comprehension, for native speakers, the effect of plausibility depended on word order. Greater plausibility effects were observed in scrambled sentences than in canonical sentences. At the same time, native speakers with higher WMC did not score better than those with lower WMC. Rather, those with lower WMC performed better on the online comprehension task. For L2 speakers, cloze scores rather than WMC influenced their accuracy. Plausibility was also found to be an important factor for their accuracy: implausible sentences were less accurately interpreted than plausible ones. In addition to the effect of plausibility, the word order effect was also reflected in their accuracy, showing higher accuracy in canonical sentences than in scrambled sentences.

4.3.5.2. Data trimming

Before online data analysis, data trimming was performed on reading times as in Experiment 1.

The excluded critical data points were 3.42% for native speakers and 3.19% for L2 speakers. When the proportion of outliers were compared by conditions, there was a significant effect of condition ($F= (3,186) = 9.126$, $p<.0001$) without a significant interaction between condition and group ($F= (3,186) = 2.098$, $p = 0.102$). Looking at the distribution of outliers across conditions (Table 4.6), it was found that more outliers were detected in the implausible scrambled condition than other conditions for both native and L2 groups. However, inspection of the distribution of outliers in the implausible scrambled condition revealed that the outliers were rather evenly distributed across items. Thus, the unequal distribution of outliers across conditions was not due to errors in materials. A possible explanation could be that difficulty of processing the implausible scrambled sentences, as shown in the accuracy rates, might results in more outliers in reading time data.

Table 4.6. Summary of reading time outliers

Group	Word Order	Plausibility					
		Plausible			Implausible		
		Number of observations	Sum of outliers	Proportion of outliers	Number of observations	Sum of outliers	Proportion of outliers
NS	Canonical	2592	63	0.024	2592	74	0.029
	Scrambled	2592	92	0.0356	2592	126	0.049
L2	Canonical	4320	127	0.029	4320	124	0.029
	Scrambled	4320	136	0.032	4320	165	0.038

Since the overall proportion of outliers was quite small and a statistical analysis using the *lmer* function in the lme4 package for R (the same analysis performed in Experiment 1) can handle unbalanced data, the detected outliers were eliminated from further analyses.

4.3.5.3. Reading time analysis

Residual reading times were computed by using the identical procedures as in Experiment 1 and saved for later analyses. Critical regions were regions 4 (scrambling effects), 5 (spill-over region), 7 (plausibility effects), and 8 (spill-over region). Same as in Experiment 1, analyses were conducted on all trials as well as on correct trials that were answered correctly in both online and offline comprehension tasks, to capture accuracy effects on reading times. From now on, analyses on all trials will be mainly discussed unless analyses on correct trials show different results.

To examine scrambling effects in regions 4 (scrambling effect) and 5 (spill-over region), the two regions were combined into one. On the analysis combining regions 4 and 5, the rationale was as follows: although the critical regions 4 and 5 were not equivalent between conditions, since one of the two regions had the same adverb (region 4 in canonical sentences vs. region 5 in scrambled sentences), if there were RT differences, it would be due to the third NP with different case-markers (NP-NOM vs. NP-ACC). For the combined two regions, an extra step was taken to prevent data distortion. Since some items did not have RTs for both regions after data trimming, such items were excluded. Thus, for regions 4 and 5, 2150 out of 2304 (93%, all trials) and 1876 out of 2304 (81%, correct trials) for native speakers, and 3594 out of 3840 (94%, all trials) and 2014 out of 3840 (52%, correct trials) for L2 speakers were included for later analyses. For region 7 (embedded verb), 1110 out of 1152 (96%, all trials) and 970 out of 1152 (84%, correct trials) for native speakers, and 1867 out of 1920 (97%, all trials) and 1045 out of 1920 (54%, correct trials) for L2 speakers were analyzed. For region 8, 1113 out of 1152 (97%, all trials) and 969 out of 1152 (84%, correct trials) for native speakers, and 1858 out of 1920 (97%, all trials) and 1044 out of 1920 (54%, correct trials) for L2 speakers were included.

Linear mixed-effects models were implemented using *lme4* (Bates, Maechler, & Bolker, 2015) for modeling and *lmerTest* package for significance testing in R (version 3.4.3; R Core Team, 2017). All modeling procedures were identical to those in Experiment 1 (a self-paced reading task) except for that working memory capacity (WMC) scores as a fixed effect were included for both native and L2 models in Experiment 2. In addition, interaction plots were constructed whenever visual inspections were required, using *effects* (version 4.0.0; Fox, 2003) and *ggplot2* (version 2.2.1; Wickham, 2009) packages.

Before discussing statistical results on residual reading times, results (raw reading times after data trimming) from native speakers are presented for visual inspection in Figure 4.4. Looking at the graph, at region 4, native speakers read scrambled sentences more slowly than canonical sentences. At region 5, a similar reading time pattern was observed. At region 7, native speakers took longer RTs in implausible sentences than in plausible sentences, and in scrambled sentences than in canonical sentences. At the spill-over region (region 8), there was a small difference in RTs between plausible and implausible sentences.

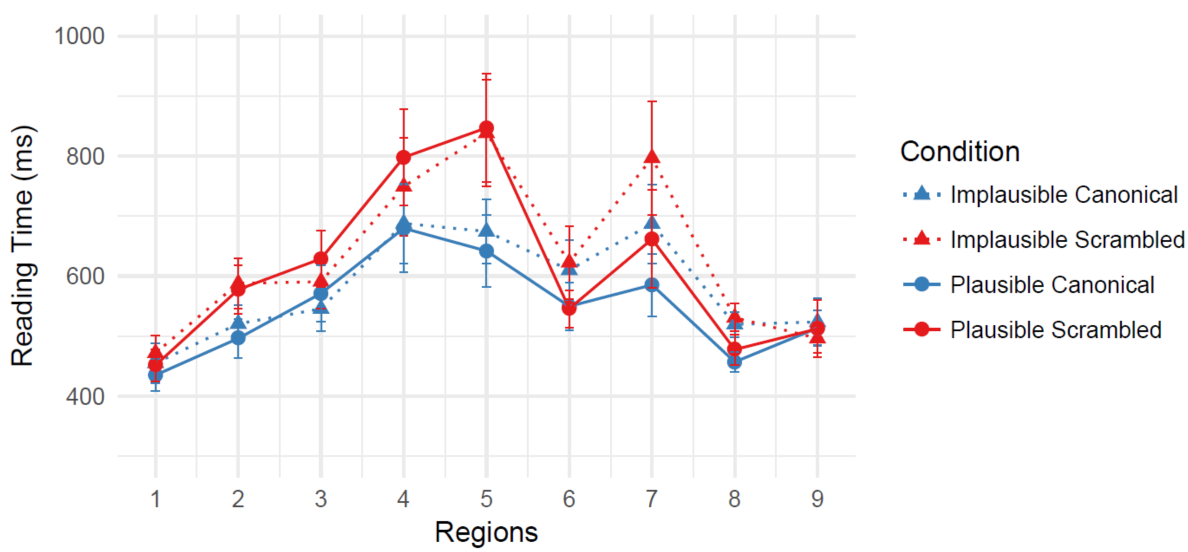

Figure 4.4. Native speakers' reading times in self-paced reading task

The results of the linear mixed effects model for regions 4 and 5 show that there was a significant main effect of word order (estimate = 269.538, SE= 86.558, $t(23.040) = 3.11$, $p= 0.005$) and a significant three-way interaction among word order, plausibility, and WMC (estimate = -49.227, SE= 22.158, $t(76.920) = -2.222$, $p= 0.029$). The significant word order effect was consistent with the RT patterns observed in Figure 4.4. (see Appendix B for the final random effects structure and a full summary of the results).

In order to understand the three-way interaction, an interaction graph was plotted as below (Figure 4.5). For plausible sentences, native speakers generally took longer RTs for scrambled sentences than canonical sentences regardless of WMC scores, while they showed different RT patterns in implausible sentences. In the implausible condition, native speakers with higher WMC took longer RTs in processing scrambled sentences than canonical sentences. Conversely, those with lower WMC (below -5) showed the reverse RT patterns. Even before the critical region (region 7), where the plausibility of sentences was disambiguated, it was found in the implausible condition that the direction of word order effects was different depending on individual differences in WMCs.

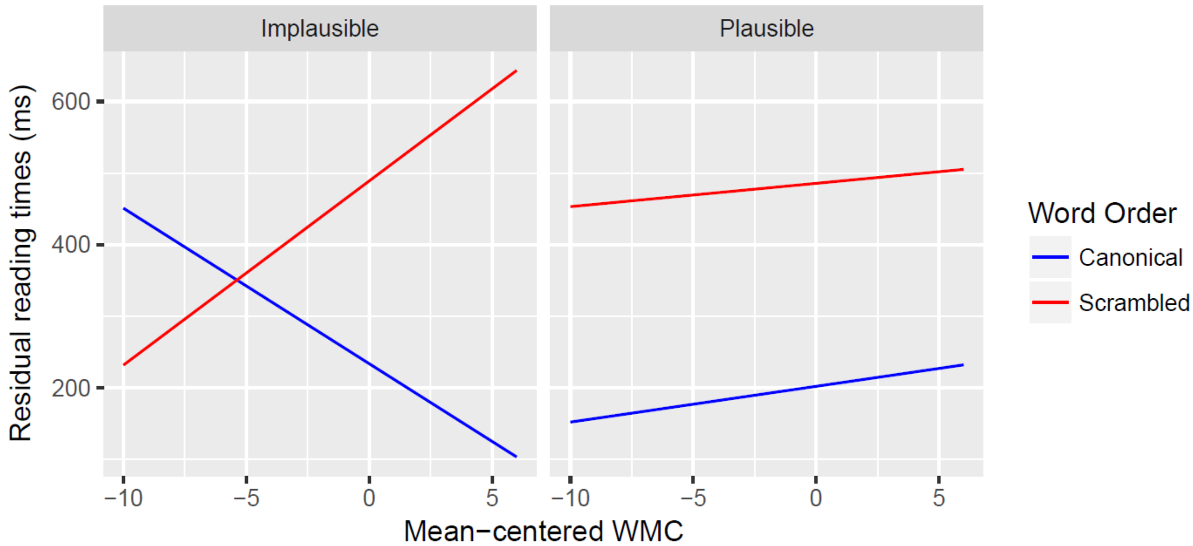

Figure 4.5. NS interaction plot among word order, plausibility, and WMC scores in regions 4 and 5: All trials

At region 7, there was a significant main effect of plausibility (estimate = -118.901, SE= 39.603, $t(45.400) = -3.002$, $p= 0.004$) with no main effects of word order (estimate = 90.740, SE= 48.482, $t(27.800) = 1.872$, $p= 0.072$) and WMC (estimate = 4.867, SE= 9.912, $t(22.500) = 0.491$, $p= 0.628$) nor significant interactions between them. Native speakers had longer RTs for implausible sentences than plausible sentences.

At region 8, a persistent main effect of plausibility was observed (estimate =-57.087, SE=16.318, $t(54.400) =-3.499$, $p=0.001$) without main effects of word order (estimate =18.389, SE=15.741, $t(60.100) =1.168$, $p=0.247$) and WMC (estimate = 7.262, SE=6.236, $t(23.700) =1.164$, $p=0.256$). There were no significant interactions between them. Same as in region 7, RTs were longer in processing implausible sentences than plausible sentences, revealing spillover effects.

Since both analyses on all trials and on correct trials in all critical regions show the same statistical results for native speakers, the results of correct trials will not be reported.

To sum up, in the critical regions for scrambling effects (regions 4 and 5), native speakers processed the scrambled sentences significantly more slowly than canonical sentences. However, at lower WMC, the effect of word order was reversed in the implausible condition. At both regions 7 and 8, native speakers demonstrated a sensitivity to the plausibility of sentences, with longer RTs for implausible sentences than plausible sentences.

L2 raw reading times are also presented in Figure 4.6 for visual inspection. At region 4, the learners had longer RTs in canonical sentences than in scrambled sentences, while at region 5, the reverse pattern was observed, showing longer RTs in scrambled sentences than in canonical sentences. At region 7, implausible sentences were processed slightly slower than plausible ones. At region 8, no clear RT differences between conditions were observed.

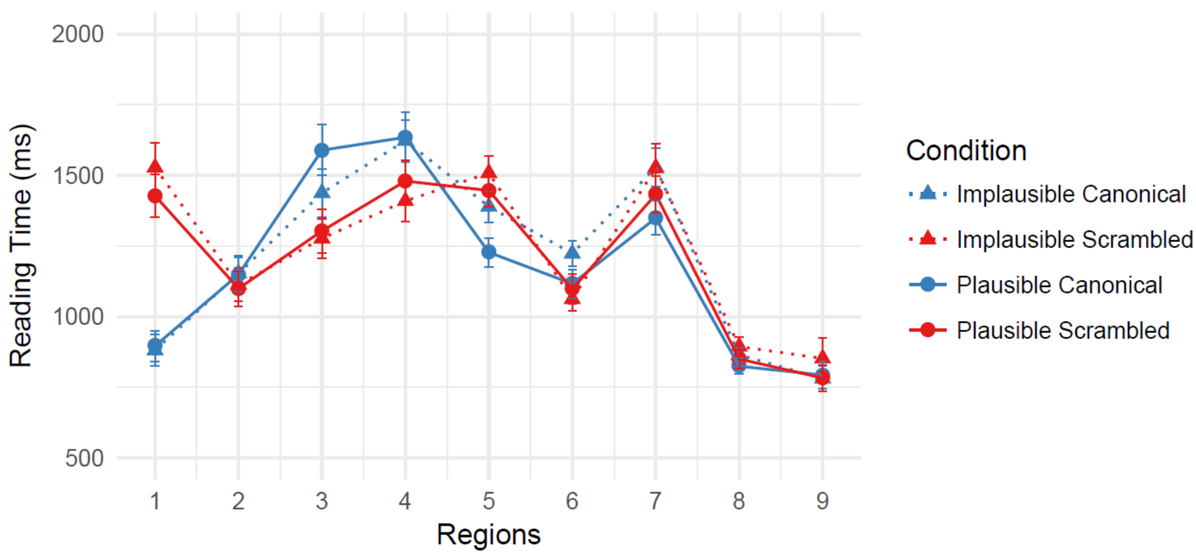

Figure 4.6. L2 speakers' reading times in self-paced reading task

At regions 4 and 5, no significant main effects of plausibility (estimate=-66.355, SE=54.664, $t(23.800)$ =-1.214, p=0.237), word order (estimate=-7.265, SE=53.750, $t(36.500)$ =-0.135, p=0.893), cloze scores (estimate=14.474, SE=16.805, $t(39.000)$ =0.861, p=0.394), or

WMC (estimate=-4.291, SE=10.250, t(35.800) =-0.419, p=0.678) were observed. However, a significant two-way interaction between word order and WMC was detected (estimate = 18.237, SE= 8.653, t(61.200) = 2.108, p= 0.039). The two-way interaction plot was built for visual inspection as in Figure 4.7. The x-axis is mean-centered WMC scores and the y-axis is residual reading times (ms). The two levels of word order are presented in two different colors.

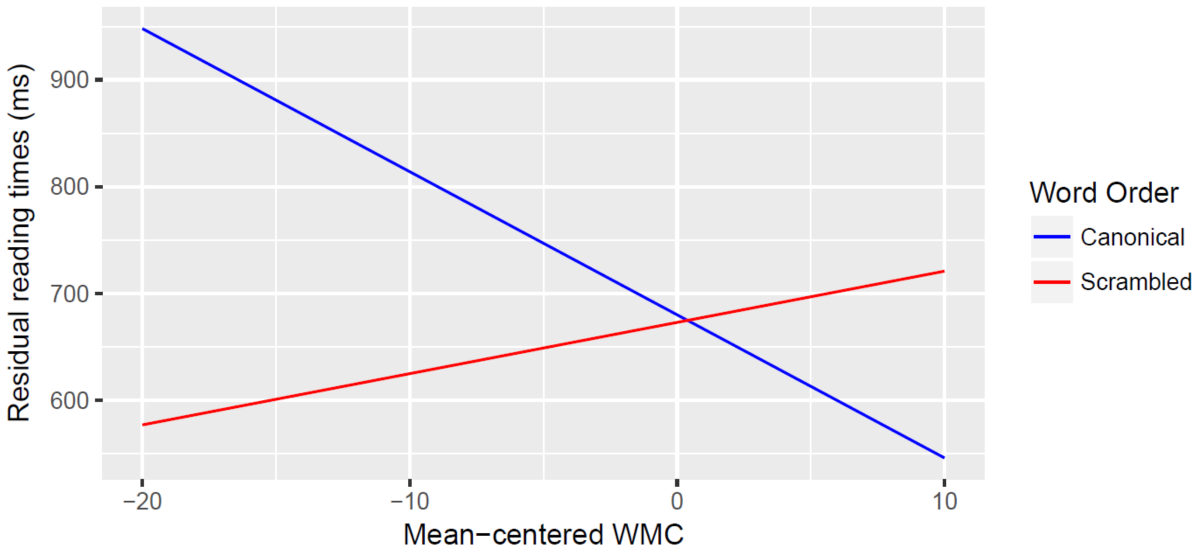

Figure 4.7. L2 interaction plot between word order and WMC in regions 4 and 5: All trials

The RT patterns were very similar to the ones found in the implausible condition with native speakers. The learners with higher WMC tended to have longer RTs in scrambled sentences than in canonical sentences, while those with lower WMC (below the average) had the reverse RT pattern. However, an analysis on correct trials did not have the same results. There were no significant main effects nor significant interactions.

At region 7, there was a significant main effect of plausibility (estimate = -134.6775, SE= 54.2833, t(36.1000) = -2.481, p= 0.0179), indicating that learners had longer RTs for implausible than for plausible sentences. A two-way interaction between plausibility and cloze scores (estimate =-33.3554, SE= 14.5350, $t(36.9000) = -2.295$, $p= 0.0275$) also reached significance, which means the effect of plausibility differed depending on learners' proficiency. The two-way interaction plot was created for a visual inspection (see Figure 4.8 below), with mean-centered cloze scores (0 equals the mean) on the x-axis and residual reading times (ms) on the y-axis.

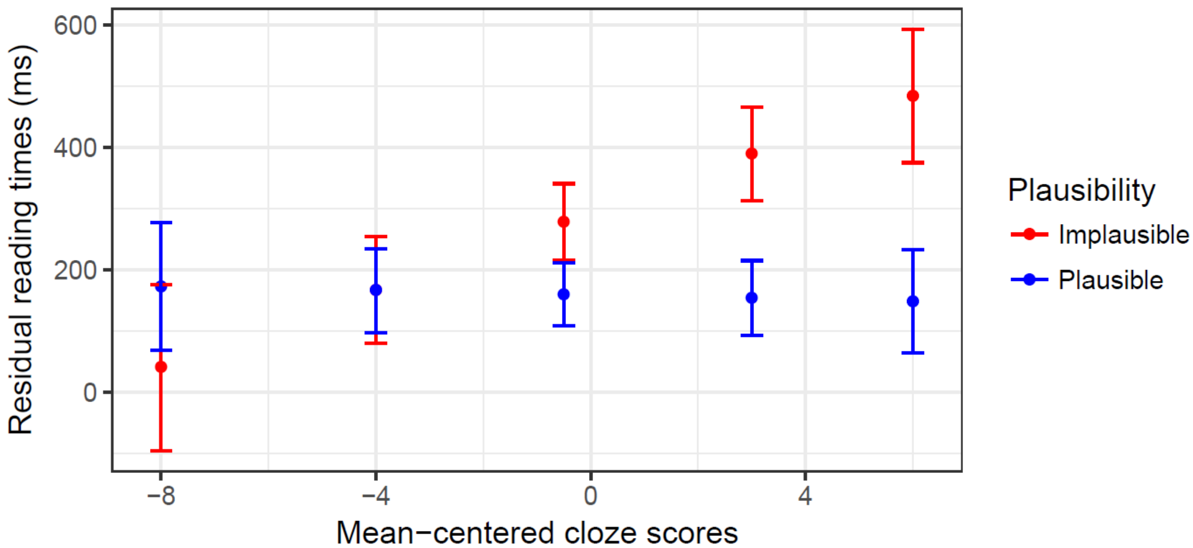

Figure 4.8. L2 interaction plot between plausibility and cloze scores in region 7 (All trials). Error bars indicate the standard error (SE).

The graph shows that L2 speakers with higher proficiency (above the mean) processed implausible sentences more slowly than plausible sentences, as predicted. On the other hand, RTs from those learners with lower proficiency (below the mean, from -4 to -8) were not different between plausible and implausible sentences, or the effect of plausibility was in the opposite direction, with longer RTs for plausible than for implausible sentences. Interestingly, the significant interaction between plausibility and cloze scores was observed in the analysis on correct trials as well (estimate = -52.491, SE= 19.971, $t(40.700) = -2.628$, $p= 0.012$). However, there was no main effect of plausibility in this analysis (estimate = -151.483, SE=77.453,

$t(41.900)$=-1.956, p= 0.057).

At the spill-over region (region 8), the main effect of plausibility was also observed (estimate =-41.051, SE=18.522, $t(39.000)$= -2.216, p=0.033) as well as a significant two-way interaction between word order and cloze scores (estimate =-11.345, SE=4.694, $t(105.700)$= -2.417, p=0.017), which was plotted in Figure 4.9. On this graph, the x-axis is mean-centered cloze scores (0 equals the mean) and the y-axis is residual reading times (ms). Negative residual reading times on the y-axis mean the reading times for this region were faster than predicted.

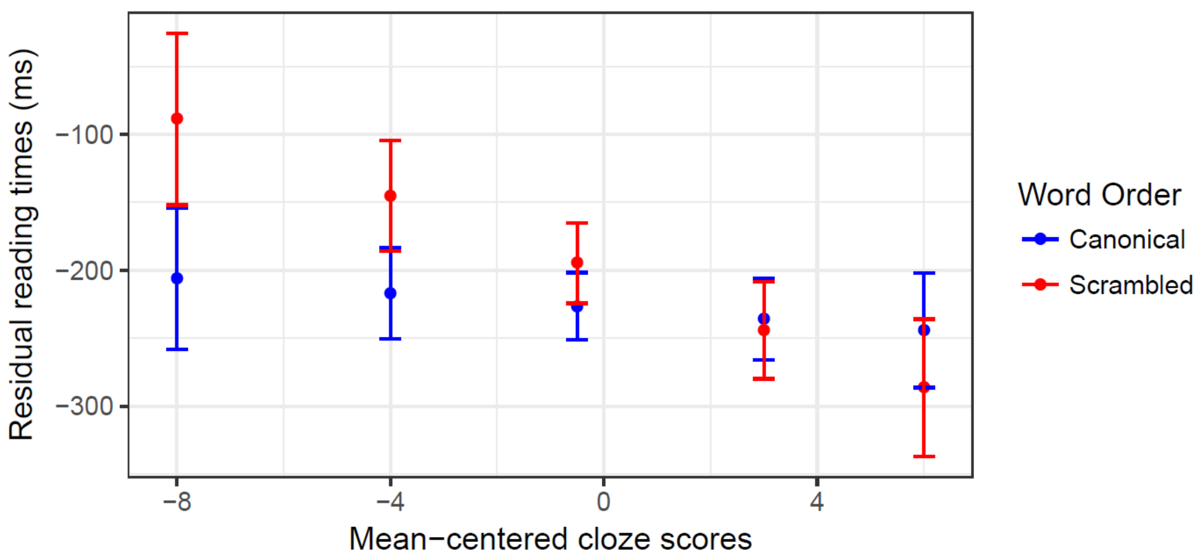

Figure 4.9. L2 interaction plot between word order and cloze scores in region 8 (All trials). Error bars indicate the standard error (SE).

The effect of word order was present only for L2 learners with lower proficiency (below the average), with longer RTs for scrambled than for canonical sentences. On the other hand, RTs from L2 learners with higher proficiency did not reveal the word order effect, either no RT differences between canonical and scrambled sentences or longer RTs for canonical than for scrambled sentences. An analysis on correct trials in this region did not show the main effect of plausibility (estimate =-62.112, SE=36.548, $t(32.800)$= -1.699, p=0.099) but the same two-way interaction between word order and cloze scores (estimate =-15.931, SE=7.068, $t(150.100)= -2.254$, $p=0.026$) was observed, indicating that even less proficient learners were trying to work out the word orders at the spill-over region.

To summarize, due to considerably low accuracy rates exhibited in online comprehension task, results of analyses on all trials and on correct trials were different in the critical regions for scrambling effects (regions 4 and 5). For all trials, RTs from L2 learners with higher WMC showed scrambling effects (longer RTs for scrambled than for canonical sentences), while the opposite RT patterns were found in L2 learners with lower WMC, with longer RTs for canonical than for scrambled sentences. No such interaction effect between word order and WMC was found in correct trials. Unlike in regions 4 and 5, the significant interaction between plausibility and proficiency (cloze scores) was observed in both analyses on all trials and correct trials in region 7. More advanced learners demonstrated a sensitivity to plausibility, with longer RTs for implausible sentences than for plausible sentences, while less advanced learners showed the reverse RT patterns. In region 8, the spill-over effect (plausibility effect) was only observed in all trials. In both all trials and correct trials, word order effects were correlated with learners' proficiency, which indicates that longer RTs for scrambled sentences than for canonical sentences were observed in less advanced learners.

4.4. Discussion

With a manipulation of plausibility, Experiment 2 investigated processing of long-distance (LD) scrambling with 24 native Korean speakers as well as 40 Mandarin-Chinese speakers. Working memory capacity (WMC) and learners' proficiency were included to measure individual differences in L2 sentence processing, same as in Experiment 1. At the same time, WMC was considered as a factor to examine differences in native sentence processing, which is different from Experiment 1.

Results from native speakers in the offline comprehension task did not show any significant effect of word order or plausibility with ceiling accuracy. For L2 speakers, proficiency was an important predictor to account for L2 offline comprehension. Highly advanced learners showed native-like performances across conditions, while less advanced learners had most difficulties integrating syntactic and semantic information in implausible scrambled sentences. Thus, the use of case-marking information varied depending on L2 proficiency when there were conflicts between plausibility information and case-marking information in scrambled sentences during offline comprehension.

Online comprehension results from native speakers showed facilitative effects of plausibility in scrambled sentences, such that they responded to scrambled sentences even more accurately than to canonical sentences when syntactic information (case-marking) was consistent with plausibility (or world knowledge). On the other hand, when it conflicted with plausibility, the reverse accuracy patterns were obtained, suggesting that plausibility information interfered with the parse based on case-marking information in implausible scrambled sentences more than in implausible canonical sentences. Individual differences in WMC were negatively associated with native online accuracy, indicating that greater WMC was disadvantageous for native speakers to perform accurately on the online comprehension task. This could be due to problem-solving strategies native speakers employed on the online comprehension task (see Beilock & Decaro, 2007 for the relation between individual differences in WMC and strategies they utilized on a math task). The native speakers with higher WMC might have taken strategies focusing on details of the given sentences rather than ‘who did what to whom’ information, which were not so effective to improve their performances on the task. On the contrary, native speakers with lower WMC might have focused on simpler problem-solving strategy (i.e., focusing on the 'who did what to whom' information) due to the limited capacity of working memory available. Further studies are needed to examine the relation between problem-solving strategies and WMC on the online comprehension task.

For L2 speakers, proficiency influenced both online and offline interpretations of sentences. For online final interpretations, case-marking information was more likely to be overridden by plausibility information regardless of word order when the two were conflicted with each other. In other words, L2 speakers tended to rely on their world knowledge when there were conflicts between the two types of information, which is different from native speakers. At the same time, although proficiency was positively related to learners' online interpretations, the overall accuracy was fairly low (see Table 4.5). Given the fact that at least highly advanced learners were able to show native-like accuracy during offline comprehensions, there was a clear discrepancy between L2 online and offline interpretations.

In order to understand what kind of information was available or prioritized during online processing before reaching final interpretations, their reading times were examined by using a self-paced reading task. Two separate analyses were conducted to address accuracy effects on reading times of all trials versus correct trials. Table 4.7. summarizes the significant main effects and interactions obtained from each group.

Table 4.7. Summary of reading time results

Group	Region	What the region shows	Significant main effects & interactions
NS	Combined regions of 4 and 5	Scrambling effects	For All trials & Correct trials: ✓ Word order effect: Canonical < Scrambled ✓ 3-way interaction among word order, plausibility, and WMC : at lower WMC, IS < IC
L2			For All trials: ✓ 2-way interaction between word order and WMC : at lower WMC, Scrambled < Canonical For Correct trials: ✓ No main effects nor interactions
NS	Region 7	Plausibility effects	For All trials & Correct trials: ✓ Plausibility effect: Plausible < Implausible
L2			For All trials: ✓ Plausibility effect: Plausible < Implausible ✓ 2-way interaction between plausibility and cloze scores : at higher proficiency levels Plausible < Implausible : at lower proficiency levels Less determinate patterns For Correct trials: ✓ 2-way interaction between plausibility and cloze scores : Same as in analysis on all trials
NS	Region 8	spillover effects	For All trials & Correct trials: ✓ Plausibility effect: Plausible < Implausible
L2			For All trials: ✓ Plausibility effect: Plausible < Implausible ✓ 2-way interaction between word order and cloze scores : at higher proficiency levels Less determinate patterns : at lower proficiency levels : Canonical < Scrambled For Correct trials: ✓ 2-way interaction between word order and cloze scores : Same as in analysis on all trials

At the regions of 4 and 5 for scrambling effects, where a filler-gap dependency is supposed to be resolved, individual differences among native speakers in WMC modulated the effect of word order with regard to plausibility. Native speakers with higher WMC showed clear scrambling effects across conditions, while those with lower WMC tended to have longer RTs for canonical sentences than for scrambled sentences in the implausible condition (RQ1). Considering that the plausibility of sentences was not disambiguated yet, the word order effects related to plausibility were not expected. A possible explanation could be that the unusual thematic relation between the embedded subject and object in the implausible condition made native speakers with lower WMC delay their completion of the filler-gap dependency. Compared to those with higher WMC, they could not afford to resolve the filler-gap relationship at the critical region (regions 4 and 5) due to limited resources available.

At the embedded verb (and a spill-over region), where the plausibility of sentences is disambiguated, all native speakers regardless of WMC showed plausibility effects, with longer RTs for implausible than for plausible sentences, suggesting they used case-marking information and resolved the filler-gap dependency (RQ2). Their successful or correct parsing was well reflected in their high accuracy rates in the online comprehension task.

For L2 speakers, at the critical regions for scrambling effects (regions 4 and 5), L2 learners with higher WMC had longer RTs for scrambled sentences than for canonical sentences, while those with lower WMC showed the reverse RT patterns when all trials were included for the analysis (RQ1 & RQ4). These RT patterns were analogous to those obtained by native speakers in the implausible condition. Applying the same explanation to the L2 RT data, the reverse RT patterns obtained by lower span L2 speakers possibly resulted from their limited resources to resolve the filler-gap dependency at the gap position, and thus delayed or failed the completion of the gap-filling. However, these effects all disappeared when only correct trials from L2 data were considered.

At the embedded verb, at least highly advanced L2 speakers showed plausibility effects, such that they had longer RTs for implausible sentences than for plausible sentences (RQ4). The same results were obtained in analyses on both all trials and correct trials. Those highly advanced learners were able to use case-marking information in real time sentence processing (RQ2). Otherwise, it is not possible to show plausibility effects since other types of information such as heuristic word order or their world knowledge cannot lead to the correct parse. The results are also related to the question raised about task effects in Experiment 1. Depending on what is required to do after reading a sentence, a probe recognition task in Experiment 1 versus a comprehension task in Experiment 2, the type of information L2 learners pay attention to or use is influenced in online sentence processing (e.g., Lim & Christianson, 2013b; 2015).

At the spill-over region in both all trials and correct trials, L2 learners with lower proficiency had longer RTs for scrambled sentences than for canonical sentences, while highly advanced learners did not. Such longer RTs in scrambled sentences at this region could indicate processing delay of the filler-gap dependency at the gap position and their attempt to complete the parsing when they encountered the verb.

So far, the online RT results are consistent with our predictions. At region 7, it is predicted that if L2 learners use case-marking information, they will show sensitivity to plausibility with longer reading times for implausible sentences than plausible sentences as follows: canonical, plausible < canonical, implausible ≤ scrambled, plausible < scrambled, implausible (see Table 4.2). Although scrambled sentences did not have longer reading times than canonical sentences (no word order effect), the predictions with respect to plausibility were borne out, providing evidence that at least L2 learners with sufficient proficiency can use case-marking information like native speakers in online sentence processing (RQ2c).

It must be also noted that the results were obtained not only from all trials but also from correct trials with much less number of data points. Then, the next question is why these RT patterns are not supported by the learners' online accuracy data if they can use case-marking information during online processing. Possibly, since the syntactic structure (i.e., LD-scrambling) tested in Experiment 2 increases processing loads, as demonstrated by reading times from native speakers in regions 4 and 5, learners could have experienced difficulties in maintaining syntactic information. In addition to the complexity of sentences, the online comprehension task at the end of sentences increases more processing loads, and thus the learners may fail to consistently rely on (or maintain) the output of syntactic parsing due to limited resources to accomplish the task.

To sum up, in Experiment 2, a clear plausibility effect was found in both native and L2 reading times, which indicate both native and advanced L2 speakers use case-marking information in online sentence processing. However, at the critical regions for scrambling effects, due to accuracy effects, word order effects were not consistently obtained in L2 data. At least in analysis of all trials, however, working memory capacity was an important predictor of L2 performances in online sentence processing. This raises another research question, the role of working memory capacity in L2 sentence processing when they process locally scrambled sentences with supportive pragmatic information. In Experiment 1, although WMC was included as a factor, the results of L2 speakers were less determinate due to task effects and native working memory capacity was not tested. Thus, in Experiment 3, both native and L2 speakers' working memory capacities will be measured as in Experiment 2 to examine the locus of L2 sentence processing difficulties.

Chapter 5

Experiment 3: Information structure and local scrambling

In Experiment 3, word order (canonical vs. local scrambled) was manipulated by using the same plausible sentences from Experiment 1. Unlike Experiments 1 and 2, plausibility was not manipulated in Experiment 3. Instead, information structure (new-given vs. given-new) was manipulated to test its interaction with syntactic structures, particularly, scrambled sentences. In addition, two factors, proficiency and working memory capacities, were examined as in Experiments 1 and 2. Experiment 3 addresses RQ1, RQ3, and RQ4.

RQ1. Does (local and long-distance) scrambling increase the processing load in L1 and/or L2 sentence processing?

RQ3. Are native and L2 speakers sensitive to information structure cues in processing canonical and scrambled sentences? If so, do information structure cues reduce the effects of scrambling in L1 and L2 sentence processing?

RQ4. Do L2 learners' proficiency and/or WMC play a role in their sentence processing? If so, is there any relationship between proficiency and working memory in L2 sentence processing?

5.1. Methodology

5.1.1. Procedure (Task Order)

Experiment 3 was conducted in a similar way to Experiments 1 and 2. The main task was a self-paced reading task, and a comprehension question followed each sentence. The comprehension question asked for 'who did what to whom' questions since it is important to make sure L2 learners can comprehend the sentences. After the online reading task, an offline acceptability judgment task was administered to check participants' sensitivity to information structures in Korean. The same cloze test, language background questionnaire, and three shortened working memory capacity tasks used in the previous experiments were given to both native and L2 speakers. Lastly, only L2 speakers were asked to complete a vocabulary task to measure their lexical knowledge of target sentences in the online reading task.

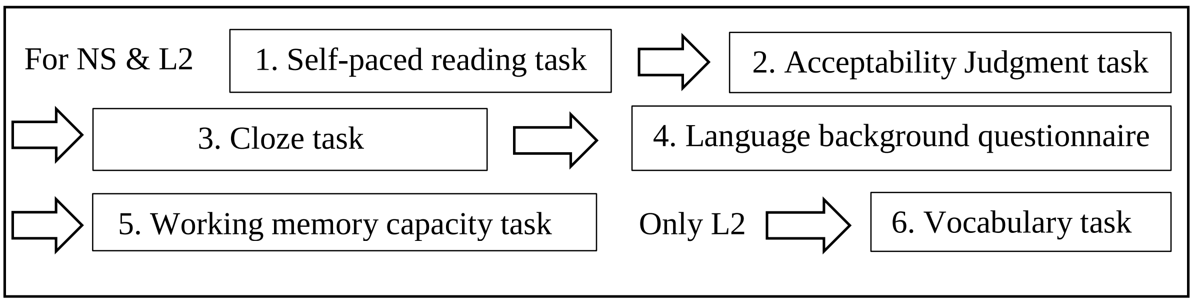

Figure 5.1. The order of tasks for Experiment 3

5.1.2. Participants

In Experiment 3, the target population was Mandarin Chinese speaking L2 learners of Korean at intermediate to advanced proficiency levels, similar proficiency levels to those in Experiment 1. Although both Experiments 1 and 3 involved locally scrambled sentences, the L2 learners in Experiment 3 were additionally given information structure cues. L2 learners with similar proficiency levels were included to examine whether they could use both syntactic (case-marking) and pragmatic (information structures) information in processing scrambled sentences. Thus, the same L2 inclusion criteria were applied to Experiment 3.

As a result, 20 native Korean speakers and 20 Mandarin Chinese speaking L2 learners of Korean were included for Experiment 3. They were all recruited and tested in Korea. Table 5.1. below summarizes language background information of the participants.

Table 5.1. Background information of participants in Experiment 3

Group	Total number	Age	First age of instruction	Length of study Korean (Months)	Length of residence in Korea (Months)	Cloze scores (Total: 40)
NS	20 (Male: 7 Female: 13	Mean: 21 Range: 19~24	N/A	N/A	N/A	Mean: 38.5 SD: 1.1 Range: 36~40
L2	20 (Male: 3 Female: 17)	Mean: 25 Range: 22~27	Mean: 20 Range: 19~24	Mean: 33 Range: 0[13]~60	Mean: 22 Range: 2~54	Mean:26.6 SD: 3.9 Range:19~32

The mean score of L2 learners on the cloze test in Experiment 1 was 26.6, same as in Experiment 3. Although the standard deviation in Experiment 1(SD = 5.3) was slightly larger than the one in Experiment 3, the two L2 groups in Experiments 1 and 3 could be considered at a similar proficiency level.

5.1.3. Materials

All the target sentences for the self-paced reading task were the same as in Experiment 1 except for that only plausible (canonical and scrambled) sentences were included in Experiment 3. Minor changes were made to the matrix subject from English names to Korean names, and an adverbial modifier appeared before an embedded verb to have a spill-over region. In addition, some kinship terms from five items used in Experiment 1 were replaced with new words as they did not work with a function word used in Experiment 3. At the same time, one subject and object of the embedded clause modified in Experiment 2 for a better interpretation was used as modified in Experiment 2.

[13] Since the length of study Korean was asking about the total periods of taking Korean classes, one participant who has not received any formal education but studied alone was considered taking no class.

In Experiment 3, instead of plausibility, information structure (new-given vs. given-new) was manipulated by pairing a context sentence and a target sentence. The new-given context was produced by using a simple affirmative sentence in which an NP was introduced, and it was the direct object of the canonical sentence and the subject of the scrambled sentence in the following target sentence. The given-new context was provided in the same way, and the noun introduced in the context was the subject of the canonical sentence and the direct object of the scrambled sentence. The four conditions are illustrated in (1) through (4).

(1) New-given context & canonical word order (New-Given-Canonical)

Context:	교실에	학생이	한 명	있었다.
	kyosil-ey	haksayng-i	han myeng	issessta
	in the classroom	student-NOM	one person	was

‘There was one student in the classroom.’

Target: 어떤 선생님이 그 학생을 조용히 가르쳤다고 민수가 생각했다.

Region 1	**Region 2**	**Region 3**	**Region 4**	**Region 5**	**Region 6**
etten	sensayngnim-i	ku	haksayng-ul	coyonghi	kaluchyessta-ko
some	teacher-NOM	that	student-ACC	silently	taught-COMP

Region 7	**Region 8**
Minswu-ka	sayngkakhayssta
Minsu-NOM	thought

‘Minsu thought that some teacher silently taught the student.’

(2) New-given context & scrambled word order (New-Given-Scrambled)

Context:	교실에	선생님이	한 명	있었다.
	kyosil-ey	sensayngnim-i	han myeng	issessta
	in the classroom	teacher-NOM	one person	was

‘There was one teacher in the classroom.’

Target: 어떤 학생을 그 선생님이 조용히 가르쳤다고 민수가 생각했다.

etten	haksayng-ul	ku	sensayngnim-i	coyonghi	kaluchyessta-ko
some	student-ACC	that	teacher-NOM	silently	taught-COMP

Minswu-ka	sayngkakhayssta
Minsu-NOM	thought

‘Minsu thought that the teacher silently taught some student.’

(3) Given-new context & canonical word order (Given-New-Canonical)

Context: 교실에 선생님이 한 명 있었다.

kyosil-ey	sensayngnim-i	han myeng	issessta
in the classroom	teacher-NOM	one person	was

‘There was one teacher in the classroom.’

Target: 그 선생님이 어떤 학생을 조용히 가르쳤다고 민수가 생각했다.

ku	sensayngnim-i	etten	haksayng-ul	coyonghi	kaluchyessta-ko
that	teacher-NOM	some	student-ACC	silently	taught

Minswu-ka	sayngkakhayssta
Minsu-NOM	thought

‘Minsu thought that the teacher silently taught some student.’

(4) Given-new context & scrambled word order (Given-New-Scrambled)

Context: 교실에 학생이 한 명 있었다.

kyosil-ey	haksayng-i	han myeng	issessta
in the classroom	student-NOM	one person	was

‘There was one student in the classroom.’

Target: 그 학생을 어떤 선생님이 조용히 가르쳤다고 민수가 생각했다.

ku	haksayng-ul	etten	sensayngnim-i	coyonghi	kaluchyessta-ko

that	student-ACC	some	teacher-NOM	silently	taught
Minswu-ka	sayngkakhayssta				
Minsu-NOM	thought				

'Minsu thought that some teacher silently taught the student.'

A context sentence always included only one NP that was modified by *one* and a classifier (e.g., *han myeong* 'one person') to introduce an entity into the discourse. As discussed in the literature review, providing one context sentence in Koizumi and Imamura (2016) and K. Jackson (2008) did not lead to different results from Park (2014) with more context sentences. Thus, it was decided to use one context sentence to reduce participants' fatigue. A target sentence had two NPs in the embedded clause and each NP was also modified by a function word, *etten* 'some' for the new entity when its entity was not mentioned in the context sentence and *ku* 'that' for the given entity when it was mentioned in the context sentence. Target sentences consisted of eight regions, and the critical regions were region 4 (second NP), where a gap for the displaced element was posited in the scrambled sentences, and region 5 (adverb) for a spill-over region, and region 6 (embedded verb) for the purpose of looking at information structure effects on the complete interpretation of the embedded clauses, although information structure effects are expected across the critical regions.

Each condition had 12 items and in total, 48 target items (4 conditions * 12 items) were included. The target items were counterbalanced across four lists, so that each participant read only one version of each set. In addition, 96 pairs of context and filler sentences with various syntactic structures were added, and the filler sentences contained six or seven regions to prevent participants from developing a strategy to predict the given or new entity in the same region. Moreover, some of the filler sentences started with one syllable function word (e.g., *i* 'this' or *ce* ‘that’) to ensure that participants got used to button-pressing function words in the main self-paced reading task. So, 20 practice items as well as 144 experimental items were presented (see Appendix A).

5.1.4. Self-paced reading task

Same as in Experiments 1 and 2, the main task was a self-paced reading task. All the procedures were the same as in Experiments 1 and 2 except for that a context sentence was presented as a whole before a target sentence. At the beginning of each trial, when a cross sign appeared on the left side of the screen, participants were instructed to press a button to read a sentence. After reading the context sentence, they were asked to press the button again, and then the second cross sign appeared on the screen. The rest of the procedures were the same as in the two experiments, reading the target sentence on a region-by-region basis.

After reading the target sentence, a comprehension question was presented as in Experiment 2, asking about the agent and the theme of the embedded clause in the target sentence. In order to distract participants from focusing on only target sentences, filler sentences had mixed questions asking about either a context sentence or a target sentence. Thus, in total, half of the experimental items had comprehension questions asking about a context sentence, and the other half asking about a target sentence. At the same time, half of the questions were designed for “yes” answers, and the other half for “no” answers.

5.1.5. Acceptability Judgment Task

Instead of an agent identification task, an acceptability judgment task was utilized in Experiment 3 to investigate whether participants were sensitive to the “given-before-new principle” in

Korean. Since the effects of information structure was examined in the online reading task, it was also tested in the offline task whether the participants would show different preferences between canonical and scrambled sentences depending on the preceding context (i.e., givenness).

The same context and target sentences were used as in the online self-paced reading task. Different from the online task, however, a context sentence was followed by both canonical and scrambled versions of the sentence. After reading the context sentence, the participants were asked to rate how natural (or preferable) each of the following sentences was on a scale from 1 (very unnatural) to 4 (very natural). They were explicitly told to indicate a preference between the two sentences, but if they did not have any preference, they were also allowed to rate both variants equally high. It should be noted that most participants tried to indicate their preferences. After the instruction, four practice items and 96 main items (48 target items+ 48 filler items) were presented, and the task was not time constrained.

In addition to the offline task, the same cloze test and language background questionnaire used in Experiments 1 and 2 were given to both native and L2 speakers.

5.1.6. Working Memory Capacity Task

The same task was given as in Experiment 1 to both native and L2 speakers (see 3.1.6).

5.1.7. Vocabulary Task

As in Experiments 1 and 2, learners' lexical knowledge was measured by using a multiple-choice vocabulary test. 66 nouns and 42 verbs used in the online self-paced reading task were included. The rest of the procedure was identical to that described in Experiment 1.

5.1.8. Predictions

The critical regions were region 4 (second NP), region 5 (adverb), and region 6 (embedded verb). Native speakers are expected to slow down at region 4 (second NP) in scrambled sentences but not in canonical sentences because of scrambling effects, as found in Experiments 1 and 2. More importantly, if a processing load associated with scrambling can be reduced, native speakers will read the scrambled sentences faster in the given-new context than in the new-given context as found in the previous literatures (e.g., K. Jackson, 2008; Kaiser & Trueswell, 2004). The effect of information structure is also expected in the canonical sentences as well due to a given-before-new advantage. However, the difference in reading time between the two contexts will not be as great as in the scrambled sentences since canonical word order sentences are less subject to information structure.

At region 5 (adverb), if there is a spill-over effect, native speakers will show the same reading time patterns. Even at region 6 (embedded verb), the same reading time patterns are expected as the second spillover region if the effect of information structure remains to influence the comprehension of the embedded clause.

For L2 learners, if they use case-marking information but fail to use information structure cues (Possibility 1), they are expected to slow down at region 4 in scrambled sentences than in canonical sentences regardless of contexts. In addition to case-marking information, if they can integrate pragmatic information (information structure) into the syntactic structure just like native speakers (Possibility 2), they would take longer to read the scrambled sentences in the new-given context than in the given-new context. The same reading time patterns are also expected in the canonical sentences with less difference in reading time between the two contexts. However, if they do not use case-marking information but rely on heuristic word order and the given-before-new principle instead (Possibility 3), they would read sentences with the given-new structure faster than with the new-given structure regardless of the word order. If they rely on heuristic word order exclusively (Possibility 4), no reading time differences are expected between canonical and scrambled sentences regardless of contexts, since a processing difficulty related to scrambling cannot be encountered without using case-marking information. Following the fourth region, the same predictions can be made at region 5 as a spillover region.

At region 6 (embedded verb), if the learners continue to rely on heuristic word order exclusively (Possibility 4), scrambled sentences would be read more slowly than canonical sentences because their inaccurate parsing leads to an interpretation where the relation between the subject and the object of the embedded verb is implausible (e.g., the student teaches a teacher). That is, a slowdown in the scrambled sentences is not due to scrambling effects but due to plausibility effects. Interestingly, it is the same reading time patterns predicted as in the case where the learners use case-marking information without integrating the information structure cues (Possibility 1). Thus, reading times in the preceding critical regions will play an important to role in teasing apart the two possible accounts of the outcome.

Furthermore, at this region, if the learners do not use case-marking information but heuristic word order and the given-before-new principle (Possibility 3), the scrambled sentences in the given-new context would be read faster than in the new-given context. Although the same reading time patterns can be predicted from the accurate parsing where the learners can integrate the information structure cues into the syntactic structure (Possibility 2), a difference should be observed in the canonical sentences. In the latter case (accurate parsing), the learners would read the canonical sentences in the given-new context faster than in the new-given context, but the difference in reading time between the two contexts should not be as great as in the scrambled sentences. On the other hand, in the former case, a significant effect of information structure is expected regardless of word order. The predictions are summarized in Table 5.2 as below.

Table 5.2. Predictions of reading time patterns at region 4 (second noun)

Predictions (qualitative similarity/difference)	Reading time patterns (A, B, C, and D refer to target sentence conditions)
Prediction 1. L1 processing = L2 processing	NS: $C < A \leq D < B$ L2: $C < A \leq D < B$ (Possibility 2)
Prediction 2. L1 processing ≠ L2 processing	NS: $C < A \leq D < B$ L2: $C = A < D = B$ (Possibility 1) OR L2: $C = D < A = B$ (Possibility 3) OR L2: $C = A = D = B$ (Possibility 4)

A: new-given, canonical; B: new-given, scrambled; C: given-new, canonical; D: given-new, scrambled

5.2. Results

5.2.1. Working memory capacity

As in Experiment 2, both native and L2 speakers completed the working memory capacity (WMC) tasks (Operation Span task (Ospan), Symmetry Span task (Sspan), and Rotation Span task (Rspan)). Data collection and scoring were identical to that in Experiment 2. Two scores, accuracy rates on distractor items and (partial) span scores on recall items, were obtained and reported for each task.

For native speakers, the average distractor accuracy was 22.9 out of 25 (92%, SD=1.9) for Ospan, and 13.2 out of 14 (94%, SD=0.8) for Sspan, and 12.7 out of 14 (91%, SD=1.3) for Rspan. For L2 speakers, the average for each task was 20 out of 25 (80%, SD=3.1), 12.2 out of 14 (87%, SD=1.7), and 11.9 out of 14 (85%, SD=1.8), respectively.

Table 5.3. below presents the descriptive statistics for the partial span scores on each task and WMC scores on the last column of the table were the sum of partial scores from the three tasks, which were used for further analyses.

Table 5.3. WMC scores of three WMC tasks

Group	Ospan (total: 25)	Sspan (total: 14)	Rspan (total: 14)	WMC (total: 53)
NS	Average: 20.5 SD: 3.5 Range: 15~25	Average: 12.2 SD: 1.7 Range: 8~14	Average: 9.9 SD: 2.7 Range: 5~14	Average: 42.5 SD: 4.5 Range: 35~48
L2	Average: 12.4 SD: 4.6 Range: 3~21	Average: 8.1 SD: 3.3 Range: 1~14	Average: 8.2 SD: 2.4 Range: 4~12	Average: 28.6 SD: 7.0 Range:12~43

5.2.2. Vocabulary task

Compared to Experiment 1, some modifications were made to the vocabulary test. Those items with poor distractors (3 items) were revised to improve the quality of the test. In addition, items that were changed in the online self-paced reading task were also replaced with the new words. Among 20 L2 speakers in Experiment 3, the average score on the vocabulary test was 100.9 out of 108 (93%, SD=3.8), which was slightly higher than 91% accuracy from L2 learners in Experiment 1. Thus, the same inclusion criterion was applied to Experiment 3, and no item was excluded.

5.2.3. Acceptability judgment task

The results of the acceptability judgment task were analyzed after converting ratings (from 1 to 4) into z-scores to factor out rater variability. The z-scores were estimated from the means and standard deviations of each subject and analyzed using *lme4* (Bates, Maechler, & Bolker, 2015) for modeling and *lmerTest* package for significance testing in R (version 3.4.3; R Core Team, 2017). Furthermore, interaction plots were constructed whenever visual inspections were required, using *effects* (version 4.0.0; Fox, 2003) and *ggplot2* (version 2.2.1; Wickham, 2009) packages.

All modeling procedures were identical to those of the offline agent identification task in Experiments 1 and 2, except for including (mean-centered) WMC scores as a continuous predictor for both native and L2 data and (contrast coded) information structure (not plausibility) as a categorical predictor. The final random effects structure for each model and a full summary of the results are presented in Appendix B.

For native speakers, there were significant main effects of information structure (estimate= -9.256e-01, SE= 1.713e-01, t(1.880e+01)= -5.403, p< .001) and word order (estimate = -6.185e-01, SE= 1.333e-01, t(1.870e+01) = -4.641, p< .001), which indicates they preferred sentences with the given-new structure to those with the new-given structure and canonical sentences to scrambled sentences. In addition, a significant two-way interaction between information structure and word order (estimate = -1.782e-01, SE= 6.216e-02, t(1.784e+03) = -2.866, p< .01) was also observed. Although there was no significant main effect of WMC (estimate = -2.942e-17, SE= 3.533e-03, t(1.784e+03) = 0.000, p= 1.000), there was a significant three-way interaction among information structure, word order, and WMC (estimate = 3.799e-02, SE= 1.438e-02, t(8.886e+02) = 2.642, p< 0.01).

For the significant two-way interaction, a visual inspection was used to identify the relation between the two factors. Figure 5.2. displays z-scores of ratings on the y-axis, two levels of information structure (given-new structure and new-given structure) on the x-axis and two levels of word order (C: canonical, S: Scrambled) in different colors.

The two-way interaction plot reveals that native speakers preferred scrambled sentences with the given-new structure to canonical sentences with the new-given structure. However, canonical sentences were preferred slightly more than scrambled sentences in the given-new condition, and for the new-given condition, the preference was much greater.

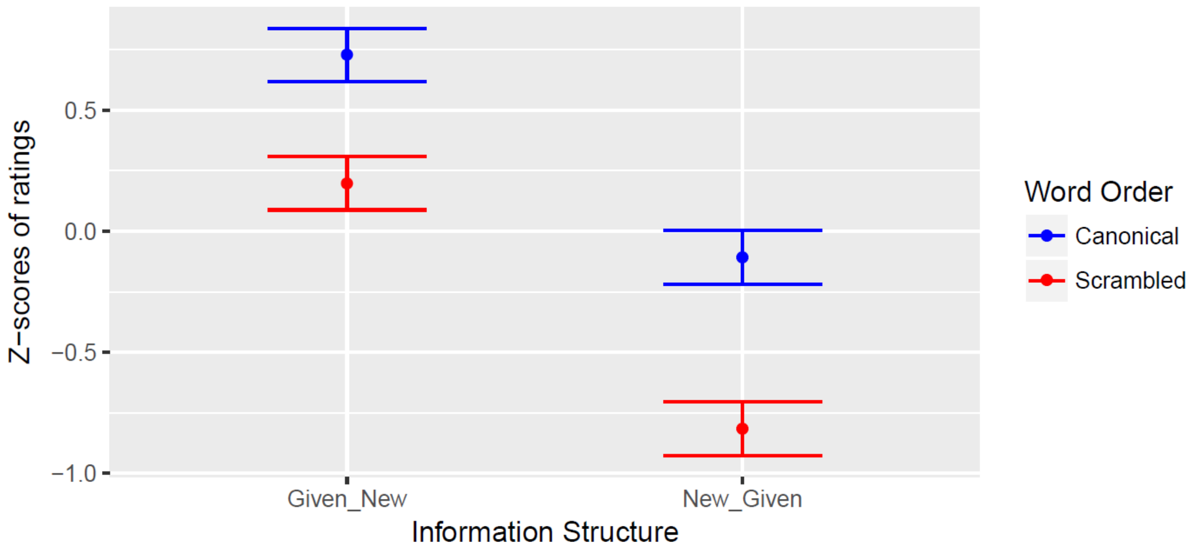

Figure 5.2. NS interaction plot between information structure and word order on AJT. Error bars represent the standard error (SE).

The three-way interaction was also plotted in Figure 5.3. Here, each panel shows the mean-centered WMC scores (0 equals the average score) and other than that, Figure 5.3. displays the same information as in Figure 5.2.

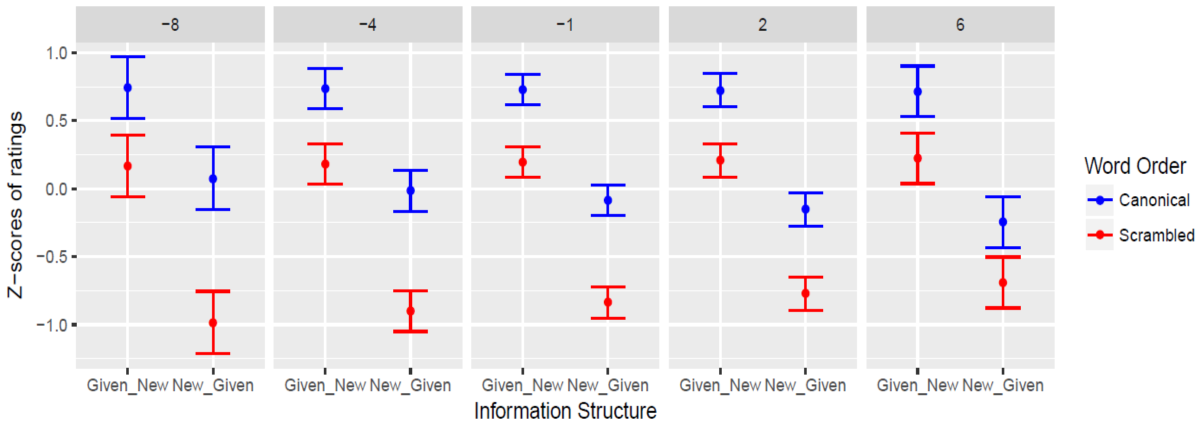

Figure 5.3. NS interaction plot among information structure and word order and WMC on AJT. Error bars represent the standard error (SE).

The overall preference pattern was like that found in Figure 5.2. Depending on the WMC, however, their preferences between canonical and scrambled sentences in the new-given context were different. For those with lower WMC, the difference was much larger than those with higher WMC.

For L2 speakers, there were significant main effects of information structure (estimate= -3.205e-01, SE= 1.465e-01, t(1.600e+01)= -2.187, p< .05) and word order (estimate = -1.059e+00, SE= 1.449e-01, t(1.710e+01) = -7.311, p< .001), indicating that L2 learners preferred sentences with the given-new structure to those with the new-given structure, and canonical sentences to scrambled sentences. No significant main effects of WMC (estimate = -4.267e-18, SE= 2.594e-03, t(1.826e+03) = 0.000, p= 1.000) or cloze scores (estimate = 1.552e-17, SE= 4.147e-03, t(1.826e+03) = 0.000, p= 1.000) or any interaction effects were observed.

To sum up, although there was an interaction between word order and information structure for native speakers, no interaction was observed for L2 speakers. That is, the effect of information structure varied between canonical and scrambled sentences for native speakers, showing a greater difference in preferences between the new-given and the given-new conditions for scrambled sentences than for canonical sentences. On the other hand, L2 speakers preferred sentences with the given-new structure over with the new-given structure, and canonical sentences over scrambled sentences.

5.2.4. Self-paced reading task

5.2.4.1. Online comprehension task

The results of the online comprehension task are presented in Table 5.4. Overall, native speakers showed high accuracy rates across four conditions, while the accuracy rates for L2 learners were lower than native speakers in all conditions. Particularly, scrambled sentences had the lowest accuracy rates regardless of information structure.

Table 5.4. Online comprehension accuracy

Group	Word Order	Information Structure			
		Given-New		New-Given	
		Proportion Correct	SD	Proportion Correct	SD
NS	Canonical	0.979	0.0370	0.983	0.0436
	Scrambled	0.979	0.0458	0.950	0.0684
L2	Canonical	0.692	0.210	0.775	0.201
	Scrambled	0.600	0.168	0.600	0.134

Statistical analyses were conducted on the accuracy rates as in Experiment 2. All the modeling procedures were identical to those in Experiment 2 except for that (contrast coded) information structure instead of plausibility was included as a fixed effect (see Appendix B for the final random effects structure for each model and a full summary of the results). Any significant interactions were plotted using *effects* (version 4.0.0; Fox, 2003) and *ggplot2* (version 2.2.1; Wickham, 2009) packages.

Native speakers demonstrated a significant two-way interaction between information structure and word order (estimate= -13.231, SE= 2.259, z value= -5.856, $p<.001$) as well as a significant three-way interaction among information structure, word order, and WMC (estimate= 0.672, SE= 0.274, z value= 2.455, $p<.05$), without any significant main effects.

For the two-way interaction, a visual inspection was performed as in Figure 5.4. The interaction plot represents, on the x-axis, the two levels of information structure and, on the y-axis, the proportional accuracy, and the two levels of word order in different line colors.

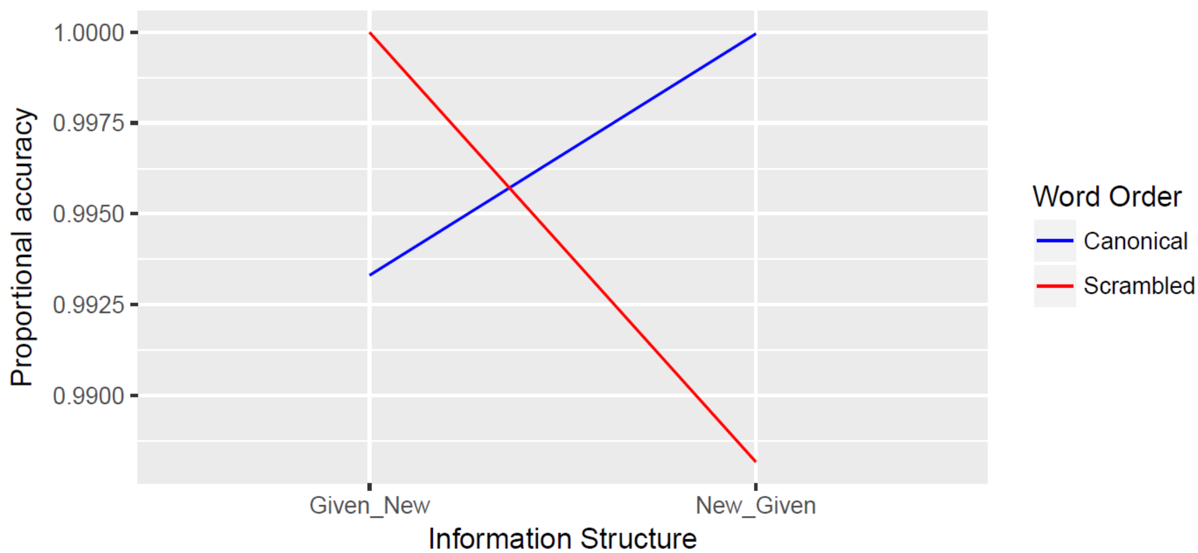

Figure 5.4. NS interaction plot between information structure and word order on the online comprehension task

Native speakers comprehended canonical sentences more accurately than scrambled sentences in the new-given condition, while the reverse accuracy pattern was observed in the given-new condition. In other words, the effect of information structure on the accuracy rates varied depending on word order.

The three-way interaction was also plotted in Figure 5.5. Each panel shows the relationship between (proportional) accuracy (on the y-axis) and the mean-centered WMC (on the x-axis) for each level of information structure (given-new and new-given).

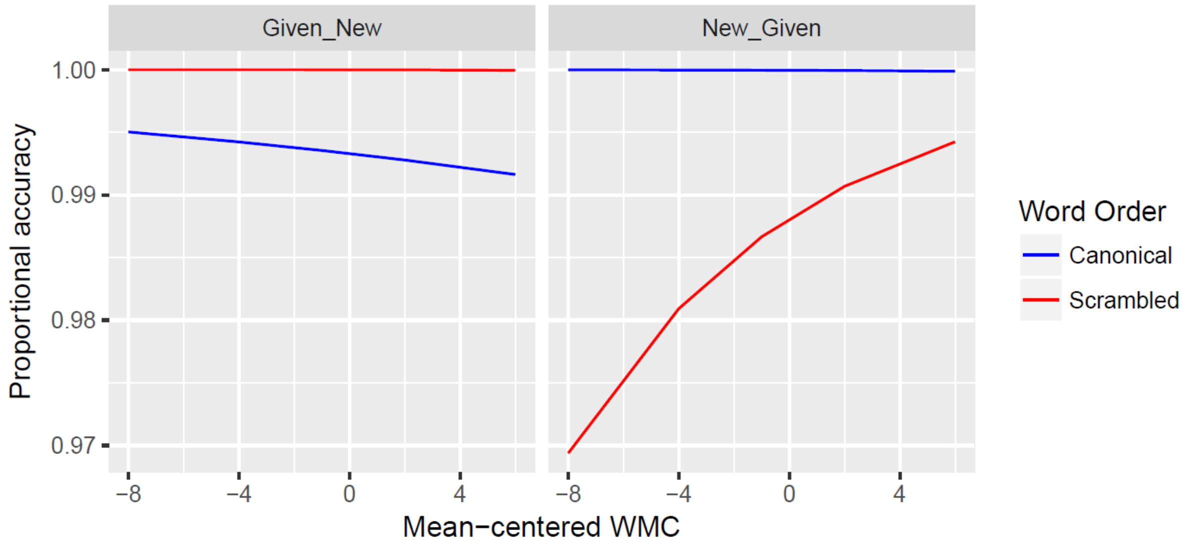

Figure 5.5. NS interaction plot among information structure, word order, and WMC on the online comprehension task

Regardless of WMC scores, accuracy rates were always higher for scrambled sentences than for canonical sentences in the given-new context, and in the new-given context, canonical sentences than scrambled sentences. The difference in accuracy between canonical and scrambled sentences was comparatively similar in the given-new context among native speakers with different WMC, while the effect of WMC was much greater for the scrambled sentences with the new-given condition. That is, native speakers with lower WMC scored much worse in scrambled sentences with the new-given structure.

For L2 speakers, there was a significant main effect of word order (estimate= -0.748, SE= 0.156, z value= -4.782, *p*<.001), indicating higher accuracy rates for canonical sentences than for scrambled sentences. In addition, a significant two-way interaction between cloze scores and WMC (estimate= -0.016, SE= 0.007, z value= -2.133, *p*<.05) was also found. The two-way interaction between the two continuous variables was plotted in Figure 5.6. The interaction plot represents the mean centered WMC on the x-axis and (proportional) accuracy on the y-axis and the mean-centered cloze scores (0 equals the average) on each panel.

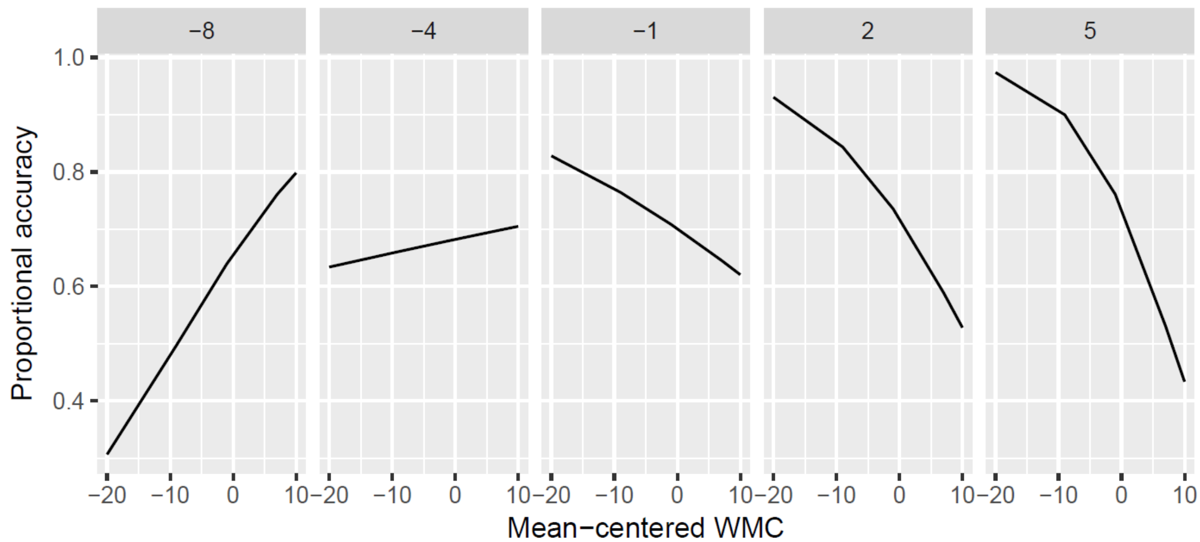

Figure 5.6. L2 interaction plot between cloze scores and WMC on the online comprehension task

For L2 learners with higher proficiency than the average, the graph showed a negative relation between the WMC and accuracy rates. In other words, the greater the WMC was, the lower the accuracy was. On the other hand, for those learners with lower proficiency (except for the L2 learners in -1 group), a positive relation between the two factors was observed. In other words, for L2 learners with higher proficiency, their accuracy rates on the task were not associated with their working memory capacity. In contrast, for lower proficient learners, individual differences in working memory capacity predicted their successful performances on the task.

5.2.4.2. Data trimming

The same data trimming procedure was used as in Experiments 1 and 2. The excluded data points accounted for 3.02% for native speakers, and 3.58% of data for L2 speakers, and there was no significant effect of condition (F= (3, 114) =3.543 , *p* =0.017) nor a significant interaction between condition and group (F= (3, 114) = 2.727, *p* = 0.047) on the outliers. Table 5.5 summarizes reading time outliers as below.

Table 5.5. Summary of reading time outliers

Group	Word Order	Information Structure					
		Given-New			New-Given		
		Number of observations	Sum of outliers	Proportion of outliers	Number of observations	Sum of outliers	Proportion of outliers
NS	Canonical	1920	55.0	0.0286	1920	43.0	0.0224
	Scrambled	1920	53.0	0.0276	1920	81.0	0.0422
L2	Canonical	1920	59.0	0.0307	1920	63.0	0.0328
	Scrambled	1920	83.0	0.0432	1920	70.0	0.0365

5.2.4.3. Reading time analysis

Residual reading times were estimated and analyzed by using the identical procedure as in Experiments 1 and 2. Statistical analyses were performed for all trials and for correct trials that were answered correctly in the online comprehension task, using *lme4* (Bates, Maechler, & Bolker, 2015) for modeling and *lmerTest* package for significance testing in R (version 3.4.3; R Core Team, 2017). Furthermore, whenever visual inspections were required, interaction plots were built using *emmeans* package, version 1.1.2 (Lenth, 2018) via *emmip* function.

The critical regions were region 4 (second NP), region 5 (adverb), and region 6 (embedded verb). Based on the online accuracy, 910 out of 960 (95%, Region 4), 907 out of 960 (94%, Region 5), 908 out of 960 (95%, Region 6) trials for native speakers, and 616 out of 960 (64%, Region 4), 621 out of 960 (65%, Region 5), 623 out of 960 (65%, Region 6) trials for L2 speakers were included for analyses on correct trials.

Linear mixed-effects models were developed, and all the modeling procedure was identical to that in Experiment 2 with an exception that information structure instead of plausibility was included as a fixed effect. Results from analyses on all trials were mainly discussed in this section, unless different results were found in analyses on correct trials (see Appendix B for the final random effects structure for each model and a full summary of the results).

Raw reading times for native speakers after data trimming are presented in Figure 5.7. According to this graph, canonical sentences were read faster than scrambled sentences regardless of information structure at region 4 (second NP). At region 5, scrambled sentences in the new-given context (New_Given_Scrambled) had longer RTs, and canonical sentences in the new-given context (New_Given_Canonical) had shorter RTs, and both canonical and scrambled sentences in the given-new context (Given_New_Canonical and Given_New_Scrambled) were between them. At region 6 (embedded verb), scrambled sentences in the new-given context had longest RTs compared to other sentences.

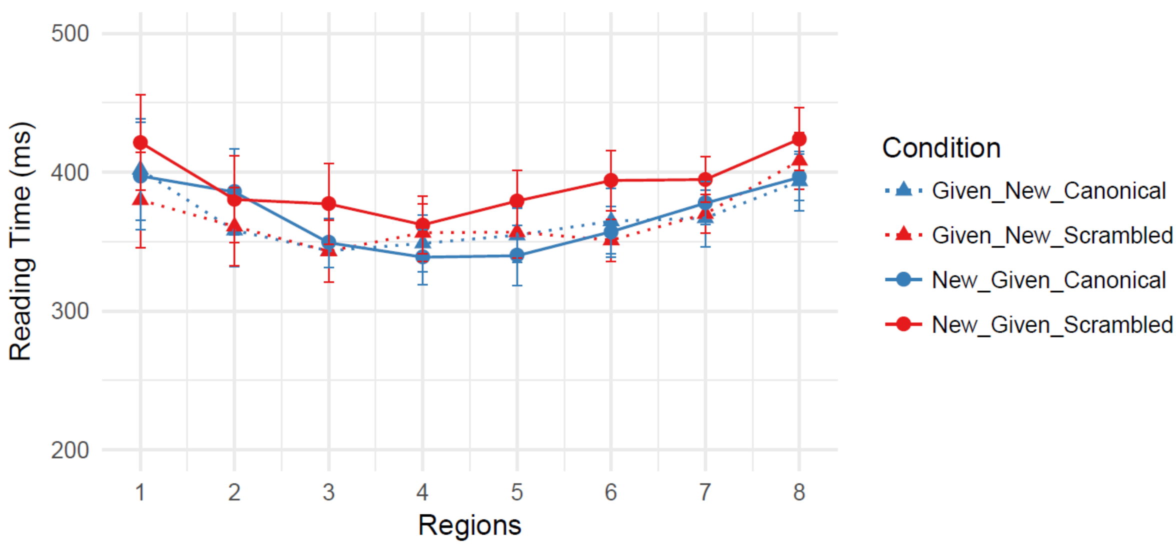

Figure 5.7. Native speakers' reading times in self-paced reading task

The results show no significant main effect of word order, information structure, WMC, nor interaction between them for analyses on both all trials and correct trials at region 4. However, at region 5, there was a significant main effect of word order (estimate = 21.004, SE= 9.873, $t(24.800) = 2.128$, $p<.05$), which indicates that native speakers read scrambled sentences more slowly than canonical sentences. Although a significant two-way interaction between word order and information structure (estimate = 37.517, SE= 15.180, $t(650.200) = 2.472$, $p<.05$) was observed, neither significant main effect of information structure (estimate = 4.430, SE= 10.801, $t(23.300) = 0.410$, $p= 0.685$), nor of WMC (estimate = 0.503, SE= 1.559, $t(17.800) = 0.323$, $p= 0.751$) were found. The significant two-way interaction was plotted in Figure 5.8. Note that the negative numbers on the y-axis means the words were read faster than predicted. Two levels of information structure (Given_New and New-Given), and of word order are presented on the x-axis and in different colors, respectively.

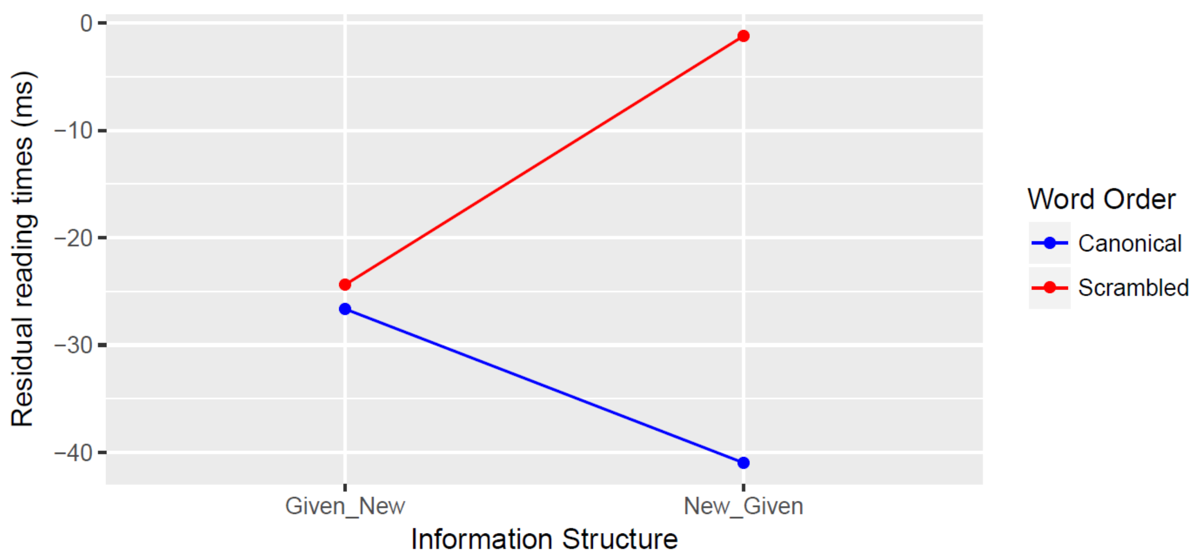

Figure 5.8. NS interaction plot between word order and information structure in region 5: All trials

In the given-new context, native speakers still took longer to read scrambled sentences than canonical sentences, but the difference in RTs was much greater in the new-given context.

Post-hoc tests using the Tukey adjusted test were performed to compare the differences in R using *emmeans* package, version 1.1.2 (Lenth, 2018). The pairwise comparisons revealed that there was a significant difference between canonical and scrambled sentences in the new-given context (p <.05), but not in the given-new context (p=0.9979). In other words, providing the given-new context facilitated processing of scrambled sentences to a greater extent.

The analysis on correct trials at region 5 did not reveal a significant main effect of word order (estimate = 20.500, SE= 10.209, t(23.800) = 2.008, p= 0.056), nor other main effects (Information Structure: estimate = 3.000, SE= 11.192, t(24.900) = 0.268, p= 0.791, WMC: estimate = 0.652, SE= 1.547, t(17.600) = 0.421, p= 0.679). However, a significant two-way interaction between information structure and word order (estimate = 36.943, SE= 15.255, t(774.900) = 2.422, p= 0.016) was also found with the same interaction pattern as above.

At region 6, there were no significant main effects of information structure (estimate = 19.859, SE= 9.691, t(23.290) = 2.049, p= 0.052), word order (estimate = 11.743, SE= 9.537, t(31.300) = 1.231, p= 0.227), or WMC (estimate = 2.190, SE= 1.255, t(22.200) = 1.745, p= 0.095), but the interaction between information structure and word order (estimate = 53.410, SE= 18.284, t(28.630) = 2.921, p<.01) was significant, which was plotted in Figure 5.9.

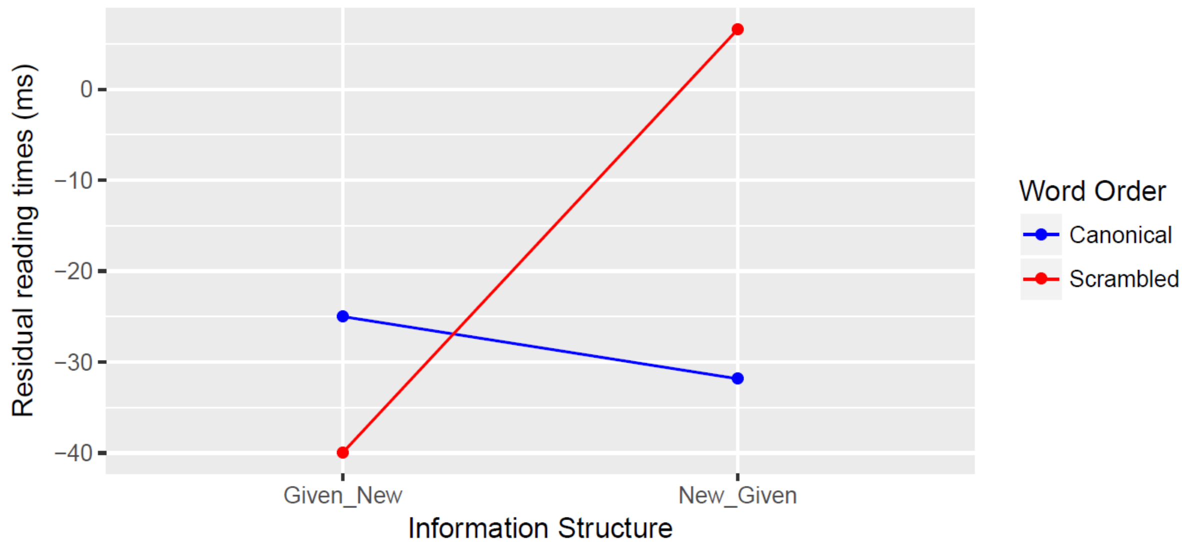

Figure 5.9. NS interaction plot between word order and information structure in region 6: All trials

Compared to Figure 5.8, the interaction pattern was slightly different. When the given-new context was provided, scrambled sentences had shorter reading times than canonical sentences. However, when the new-given context was presented, scrambled sentences took longer reading times than canonical sentences. Again, post-hoc tests were performed in the same way as in region 5, and the results showed that there was a significant difference between canonical and scrambled sentences in the new-given context (p <.05), but not in the given-new context (p = 0.7320). In addition, a significant difference between scrambled sentences in the new-given context and in the given-new context was detected (p <.001). Thus, the facilitative effect of information structure on the processing of scrambled sentences was clearly manifest in the given-new condition at region 6.

The same results were found in the analysis on correct trials, a significant two-way interaction between information structure and word order (estimate = 53.749, SE= 18.764, t (27.660) = 2.865, p<.01).

To sum up, in the new-given context, scrambled sentences showed longer reading times than canonical sentences at the critical regions 5 and 6. However, the facilitative effect of information structure was observed in the two regions on the processing of scrambled sentences in the given-new context. At region 6, scrambled sentences were even read faster than canonical sentences in the given-new context.

L2 raw reading times after data trimming are shown in Figure 5.10 as below.

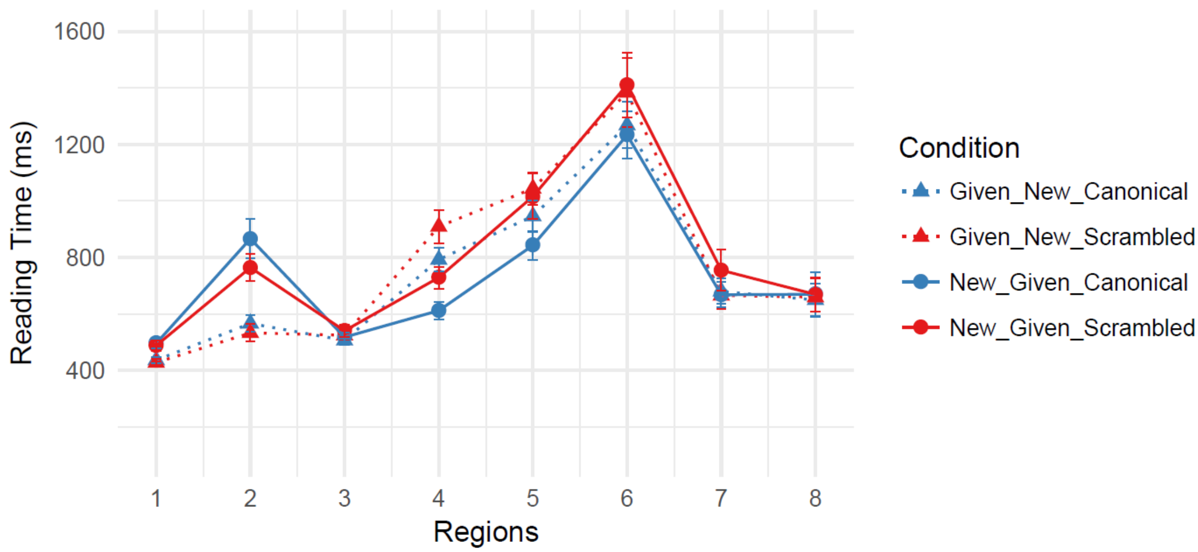

Figure 5.10. L2 speakers' reading times in self-paced reading task

Across critical regions (regions 4, 5, 6), canonical sentences in the new-given context (New_Given_Canonical) were read fastest, which was also found in native reading times in Figure 5.7. Different from native speakers, however, L2 speakers tended to read scrambled sentences in the given-new context (Given_New_Scrambled) more slowly than sentences in other conditions.

At region 4, statistical results reveal significant main effects of information structure (estimate =-187.447, SE=31.411, $t(20.500)$ =-5.968, p<.001) and word order (estimate =63.160, SE=30.462, $t(40.400)$ =2.073, p=<.05), but no other main effects (Cloze_scores: estimate =

1.301, SE= 6.329, $t(21.200) = 0.206$, $p= 0.839$, WMC: estimate = 0.538, SE= 3.627, $t(15.900) = 0.148$, $p= 0.884$), nor interaction effects were found. That is, new-given sentences were read faster than given-new sentences, and canonical sentences faster than scrambled sentences. The analysis on correct trials shows only a significant main effect of information structure (estimate =-171.417, SE=39.256, $t(19.400) =-4.367$, $p<.001$) without other main effects. Thus, the effect of information structure was also found in the analysis on correct trials, but a word order effect (or scrambling effect) was not present.

At region 5, there were significant main effects of information structure (estimate= -76.041, SE= 32.201, $t(17.800)= -2.361$, $p= 0.030$), word order (estimate = 129.016, SE= 30.157, $t(39.800) = 4.278$, $p<.001$), and cloze scores (estimate = -17.575, SE= 7.979, $t(18.000) = -2.203$, $p= 0.041$), such that new-given sentences were read faster than given-new sentences, and canonical sentences faster than scrambled sentences. At the same time, more proficient L2 learners read sentences faster than learners with lower proficiency. No significant interactions were found. Without the main effect of cloze scores (estimate = -15.955, SE= 8.080, $t(16.200) = -1.975$, $p= 0.066$), the analysis on correct trials also reveals significant main effects of information structure (estimate = -78.756, SE= 38.265, $t(66.600) = -2.058$, $p= 0.043$) and word order (estimate = 128.416, SE= 36.260, $t(62.300) = 3.542$, $p<.001$). Thus, in both analyses, L2 speakers read scrambled sentences more slowly than canonical sentences regardless of information structure. In addition, they read new-given sentences faster than given-new sentences. Different from native speakers, no interaction between word order and information structure was found.

At region 6, a significant main effect of word order (estimate = 135.827, SE= 55.253, $t(29.900) = 2.458$, $p<.05$) was observed, such that scrambled sentences were read more slowly than canonical sentences. At the same time, more proficient learners read sentences faster than lower proficiency learners (Cloze scores: estimate = -31.554, SE= 13.942, $t(21.300) = -2.263$, $p<.05$). No other main effects, nor interaction effect were found. On the other hand, the analysis on correct trials reveals a significant four-way interaction among information structure, word order, cloze scores, and WMC (estimate = -9.655, SE= 4.070, $t(545.100) = -2.372$, $p<.05$) without other main effects nor interactions. The four-way interaction was plotted using *effects* (version 4.0.0; Fox, 2003) and *ggplot2* (version 2.2.1; Wickham, 2009) packages as below (Figure 5.11). On the x-axis is the mean-centered WMC (0 equals the average) and on the y-axis is the residual reading times (ms). Different line colors correspond to the two levels of word order. Solid lines indicate given-new sentences, and dotted lines indicate new-given sentences. Each panel shows the mean centered cloze scores.

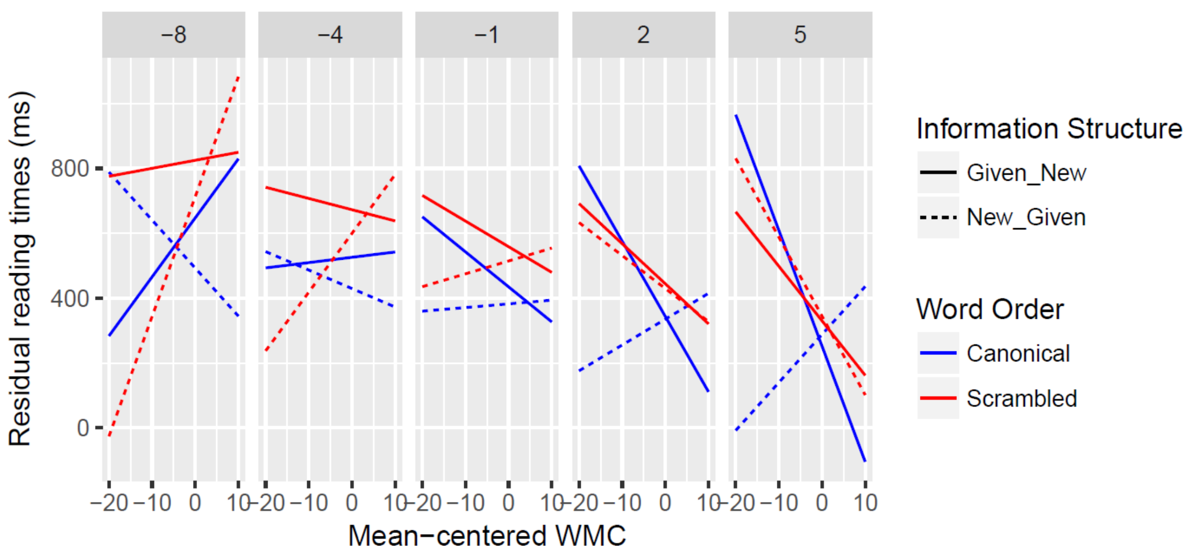

Figure 5.11. L2 4-way interaction plot among information structure, word order, WMC, and cloze scores: Correct trials

Since one of the research questions was about the relationship between proficiency and WMC, it was decided to focus on L2 learners with greater WMC (between 5 and 10, above the average) to test whether learners with high WMC could perform more like native speakers. According to the graph, for learners with lower proficiency (below the average), scrambled sentences in the new-given context were read more slowly than in the given-new context. However, unlike native speakers, they read scrambled sentences more slowly than canonical sentences regardless of information structure. In other words, the interaction between word order and information structure was not observed. For learners with higher proficiency (above the average), the effect of information structure was much greater for canonical sentences than for scrambled sentences. That is, canonical sentences in the new-given context were read much more slowly than in the given-new context, while there was little reading time difference between scrambled sentences in the two conditions.

In sum, the results from analyses on both all and correct trials show that the given-new sentences were read more slowly than the new-given sentences regardless of word order at region 4. It was at region 5 that the word order effect was observed in addition to the effect of information structure. At this region, L2 learners read scrambled sentences more slowly than canonical sentences, and the given-new sentences more slowly than the new-given sentences. At region 6 (embedded verb), when all trials were analyzed, it was found that L2 learners read scrambled sentences more slowly than canonical sentences, and that more proficient learners tended to read faster than less proficient learners. The results from the analysis on correct trials revealed different reading time patterns depending on individual differences in proficiency and WMC.

5.3. Discussion

In Experiment 3, processing of locally scrambled sentences with different information structures was examined. Compared to Experiment 1, the main research question was whether L2 learners could use and integrate case-marking information and pragmatic cues (information structure) during processing of scrambled sentences.

For the offline acceptability judgment task, the results showed that native speakers were sensitive to information structure in Korean as claimed in the previous literature, and their preferences for scrambled sentences were greatly influenced by information structure. They generally preferred canonical sentences to scrambled sentences when the sentences had the same information structure, but they preferred the scrambled sentences in the given-new condition to canonical sentences in the new-given condition. In other words, there was an interaction between information structure and word order. Furthermore, individual differences in working memory capacity also influenced their preferences, such that native speakers with lower WMC tended to have stronger preferences of canonical sentences to scrambled sentences in the new-given condition. Without the given-before-new advantage, the native speakers with lower working memory might have more difficulty in maintaining new information in the new-given condition, and such difficulty could lead to their lower acceptance of non-canonical (scrambled) sentences, which requires more complex parsing. In the same offline task, L2 speakers also preferred given-new sentences to new-given sentences, and canonical sentences to scrambled sentences. However, there was no interaction between the two factors. Thus, differences between native and L2 speakers were found in the extent to which they accepted scrambled sentences in the given-new condition compared to canonical sentences in the new-given condition.

Such preference patterns were reflected in online reading time results from native speakers. A significant interaction between information structure and word order was observed at region 5 (adverb, a spill-over region) in analyses on both all and correct trials, such that native speakers read scrambled sentences more slowly than canonical sentences in the new-given condition (scrambling effects), while little difference in reading times was observed in the given-new condition. Likewise, a significant interaction between the two factors was also observed at region 6 (embedded verb) when they completed an interpretation of the embedded clause. Across critical regions, native speakers' sensitivity to information structure with respect to word order was reflected in their online sentence processing: in the new-given condition, scrambled sentences still had longer reading times than canonical sentences, while such processing difficulty was reduced in the given-new condition (RQ1 and RQ3).

For L2 learners, since their online accuracy was lower than native speakers, there was a discrepancy between results of analyses on all trials and on correct trials. At region 4, the effect of information structure was present in both analyses, but the word order effect disappeared in the analysis on correct trials. At region 5 (a spill-over region), regardless of online accuracy, both effects of word order and information structure were found without an interaction between them. That is, L2 speakers read scrambled sentences more slowly than canonical sentences regardless of information structure, and given-new sentences more slowly than new-given sentences regardless of word order. It is important to note that L2 learners in Experiment 3 showed scrambling effects in online processing, which was not found in Experiment 1 with L2 learners of similar proficiency levels (RQ1). The results provide evidence of task effects which we suggested as a possible account of L2 results (no scrambling effects) in Experiment 1. The discrepancy between the presence and absence of scrambling effects from L2 data in Experiments 1 and 3 could be explained by the nature of tasks (cf. Lim & Christianson, 2015), in that comprehension questions in Experiment 3 required participants to complete their interpretations by paying attention to case-marking to answer the questions, while L2 learners in

Experiment 1 did not have to do so to complete a probe recognition task. By comparing L2 learners with similar proficiency levels in Experiments 1 and 3, it is possible to claim that L2 learners in Experiment 1 were also capable of using case-marking information in online sentence processing.

Unlike case-marking information, L2 learners' sensitivity to information structure was not reflected in their reading times like native speakers (RQ3). Rather, the results seemed to be consistent with our prediction (Possibility 1), where L2 learners use case-marking information without integration of information structure during processing of sentences. Given the fact that given-new sentences had longer reading times than new-given sentences at regions 4 and 5, even though the L2 speakers also showed the given-before-new preference in the offline task, the online results seemed to indicate that L2 learners did not actually show effects of information structure but lexical priming effects. Shorter reading times at region 4 (as well as 5) in the new-given sentences than in the given-new sentences could be attributed to the occurrence of the same noun that was given in a context sentence. Hence, it appears that L2 speakers in Experiment 3 could use case-marking information but did not use information structure as natives did, and in turn could not integrate information structure during online processing.

Finally, among the three critical regions, the effect of online accuracy was most apparent at region 6 (embedded verb). In analysis of all trials, significant main effects of word order and cloze scores (proficiency) were observed, indicating that L2 learners processed scrambled sentences more slowly than canonical sentences as in the previous regions, and proficiency influenced their reading times. When only correct trials were analyzed, a significant interaction among cloze scores (proficiency), WMC, word order, and information structure was detected. Since our last research question was about the relationship between proficiency and WMC, the interaction was closely examined. Results from L2 learners with higher WMC showed different reading time patterns depending on their proficiency. L2 learners with lower proficiency processed scrambled sentences more slowly than canonical sentences, particularly, scrambled sentences in the new-given than in the given-new conditions, with no interaction between word order and information structure. Thus, at least for scrambled sentences, the L2 learners with lower proficiency with greater WMC showed the given-before-new advantages. On the other hand, for L2 learners with higher proficiency with greater WMC, information structure influenced processing of canonical sentences rather than scrambled sentences, in that they processed canonical sentences in the given-new condition much faster than in the new-given condition, while scrambled sentences in the two types of information structures showed little difference. This is a completely opposite interaction pattern. Although native speakers had the facilitative information structure effect (i.e., the given-before-new advantage) in processing scrambled sentences, L2 learners with higher proficiency with greater WMC had the same effect when they processed canonical sentences. The results suggest that native speakers are likely to associate information structure with scrambled sentences, while L2 learners with higher proficiency with greater WMC tend to do so with canonical word order sentences.

Therefore, according to the results from Experiment 3, it is found that L2 learners can pay attention to and use case-marking information during online sentence processing if a given task requires complete interpretations. In addition, both native and L2 speakers showed scrambling effects in the new-given context, supporting for our claim that processing of scrambled sentences is associated with more processing due to a syntactic complexity. However, native and L2 speakers differ from each other regarding use of information structure during online processing. Native speakers tend to associate information structure with scrambled sentences, while L2 learners even with higher proficiency with greater WMC integrate that information with canonical sentences. In this respect, it can be claimed that the locus of difficulties in L2 sentence processing could be due to the interface between syntax and pragmatics.

Chapter 6

Discussion and conclusion

6.1. General Discussion

The present study sets out to investigate whether and how L2 learners of Korean utilize morphosyntactic (case-marking) or non-syntactic (plausibility, heuristic word order, or information structure) information in processing locally or long-distance scrambled sentences as compared to native Korean speakers, and whether individual factors such as proficiency and working memory capacities play a role – in particular, whether advanced learners are able to attain native-like use of morphosyntactic information.

For L1 sentence processing, this study is designed to provide evidence bearing on whether scrambling involves an additional processing load due to the postulation of gap. In Experiment 1, native Korean speakers showed scrambling effects in online sentence processing. Their reading times at the gap position (i.e., at the embedded verb) were longer for scrambled sentences than canonical word order sentences. For long-distance (LD) scrambling in Experiment 2, an additional processing loads, as predicted, were observed at the gap position (i.e., the combined region of the third NP and the adverb).

In Experiment 1, however, since the processing cost showed up on the embedded verb, it could be argued if the cost is associated with thematic role assignment once the verb is encountered (The Direct Association account: Pickering & Barry, 1991; Pickering, 1993, 1994). Due to the adjacency between the second NP and the verb, results from Experiment 1 cannot tease apart the two accounts (gap-based account versus the direct association account), but Experiment 3 can provide relevant evidence against the direct association account. Scrambled sentences in Experiment 3 have a similar syntactic structure with scrambled sentences in Experiment 1 except for that their information structures are manipulated by a preceding context sentence. If scrambled sentences in new-given context are considered, however, as the preceding discourse context in this condition does not give any clues about the referent of the scrambled NP, the sentences are quite comparable to the scrambled sentences without a preceding context in Experiment 1. In Experiment 3, locally scrambled sentences have an intervening adverb between the second NP (a gap is assumed to be postulated) and the embedded verb. That is, the gap is not adjacent to the verb unlike in Experiment 1. If the slowdown found at the verb in Experiment 1 is attributed to the verb, we cannot expect to observe any slowdown effect (or processing difficulties) at the adverb prior to the verb in Experiment 3. The results in Experiment 3 instead reveal a significant scrambling effect at the adverb in the new-given context, showing longer reading times in scrambled sentences than in canonical sentences at the adverb (presumably as the gap-posited position or as the post-gap position depending on when the gap is postulated) before the subcategorizing verb. In this respect, the slowdown observed at the verb in Experiment 1 is highly likely to be associated with the postulation of gap rather than the verb, which theoretically corresponds to the movement-based account of scrambling (filler-gap dependencies).

For L2 sentence processing, the major finding of this study is the effects of individual factors such as proficiency and working memory capacities. In Experiment 1, both online and offline results from L2 learners show clear proficiency effects. For the offline agent identification task, more proficient L2 learners were likely to score better. At the same time, the offline results show that plausibility had a facilitative effect on the comprehension of scrambled sentences. Unlike native speakers, however, L2 data on the online sentence processing yielded mixed results depending on L2 proficiency. At the critical region of the second NP, more advanced L2 learners showed scrambling effects, while less advanced learners showed less determinate reading time patterns. However, at the embedded verb where the plausibility of sentences should be obvious, even those more advanced learners did not process sentences in a native-like manner, instead showing their reliance on heuristic word order and plausibility information, and restricted use of case marking information in processing scrambled sentences.

For learners' inconsistent use of morphosyntactic information in processing scrambled sentences, task effects are suggested as a potential modulating factor. In previous literature, for example, in Williams (2006), similar task effects are also observed. When a given task required participants to pay more attention to plausibility like judging plausibility at each word in the stop-making-sense task, L2 participants showed sensitivity to plausibility like native speakers. However, when such requirement was removed, and they needed to read sentences to answer memory questions, the plausibility effect was no longer observed at the same region. Likewise, Lim and Christianson (2013b) also provide strong evidence of task effects on L2 processing by manipulating reading goals (for translation versus for comprehension), suggesting that L2 learners may perform differently depending on what is required after reading.

In this study, although we had a clear reason not to include comprehension questions (i.e., not to make the task too explicit) in Experiment 1, since the possibility of task effects need to be considered in interpreting L2 results, learners' comprehension of sentences in online tasks in Experiments 2 and 3 were measured. In Experiments 2 and 3 where comprehension questions essentially make them focus on case-marking information to answer questions, clear scrambling effects were observed with effects of individual factors. Methodologically, it seems important to consider the potential modulating effects of task when complex sentences are tested. Without task demands, learners or even native speakers may not pay enough attention to syntactic information, leading to difficulties to interpret results.

In terms of processing of LD scrambling in Experiment 2, a significant interaction between word order and working memory capacities was observed at the gap position (i.e., the combined region of the third NP and the adverb). That is, L2 learners with high working memory capacities processed the LD scrambling in a native-like manner, while those with lower working memory capacities failed to do so. As discussed in L1 sentence processing, the slowdown observed at this combined region represents processing difficulties associated with the postulation of gap before encountering the embedded (subcategorizing) verb. For such native-like processing of sentences with LD scrambling, it is assumed that learners should be able to use not only case marking information, but also be able to maintain and retrieve the scrambled NP from their working memory. Thus, the correlation between word order and working memory indicates that working memory capacity is indeed a crucial predictor for native-like processing of LD scrambling.

Interestingly, the effect of working memory is also observed at this region among native speakers. Native speakers with lower WMC processed implausible canonical sentences more slowly than implausible scrambled sentences. Compared to L2 learners with lower WMC who had longer reading times for canonical sentences than scrambled sentences, those native speakers with lower WMC seem to have difficulty in making sense of implausible sentences even prior to the embedded verb (i.e., the disambiguating region for the plausibility). It has been well documented in the recent literature that native speakers can anticipate upcoming information as information unfolds (Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003; Van Berkum et al., 2005; see Kaan, 2014 for more discussion about L1 and L2 predictive processing).

Thus, it is possible for native speakers in this study to predict the plausible relation between the three NPs as they encounter each NP, but when syntactic encoding and their predictions don't align, processing difficulties occur among those native speakers who have less cognitive resources to handle such mismatches as well as to resolve the filler-gap relationship. In this respect, the observation that even native speakers with lower WMC differ from those with high WMC in processing the filler-gap dependency favors the processing capacity approach (PCA), in that the same individual factor, working memory, also results in different processing patterns among native speakers, as found in L2 processing.

In addition to the effect of working memory capacity at the gap position in Experiment 2, a significant interaction between plausibility and proficiency was also detected at the embedded verb. L2 learners with advanced proficiency showed plausibility effects like native speakers, indicating that they could successfully use case-marking information. On the other hand, learners with lower proficiency showed less determinate patterns.

As outlined in chapter 2, two approaches, the SSH (Shallow Structure Hypothesis) and the PCA (Processing Capacity Approach), have made different predictions about L2 processing in Experiment 2. The SSH hypothesizes that all learners rely exclusively on non-syntactic cues regardless of whether they are intermediate or advanced L2-learners. So, if the SSH is correct, none of the learners in Experiments 2 should use morphosyntactic information for incremental processing or have difficulty integrating case-marking information. On the other hand, if the PCA is correct, more advanced learners or learners with high working memory capacities should be able to integrate case-marking (morphosyntactic) information and process sentences in a target-like manner. The results in Experiment 2 support the PCA rather than the SSH.

However, the SSH could argue that it is possible to analyze case markers as cues for thematic information rather than syntactic information. That is, learners' use of case marking information can be understood as their sensitivity to thematic cues and thus does not necessarily indicate that they can have a full syntactic representation like native speakers. An argument can be compatible with the results found at the embedded verb. Learners may have had less difficulties assigning the thematic roles on each NP (e.g., Nominative as Agent, Accusative as Patient) as reanalyzing case marking information as thematic cues. However, the real challenge for the SSH is the resemblance between native and L2 speakers found at the gap position, the slowdown associated with resolving the long-distance filler-gap dependency. According to the SSH, L2 learners can only build syntactic construction lacking such syntactic details. Although results from the learners with lower WMC confirm their predictions, crucially, similar performances are also found among native speakers with lower WMC. Unless the SSH can account for the similarities found at the gap position between L1 and L2 processing, it can be claimed that the results in Experiment 2 are more compatible with the PCA, attributing the L1 and L2 differences to processing difficulties rather than to representational deficits posited by the SSH.

Different from Experiments 1 and 2, an interaction between information structure and word order is examined in Experiment 3 by using locally scrambled sentences. For L1 processing, the significant interaction between the two factors was observed in the offline acceptability judgment task, and such interaction effect was well reflected in their reading time data: native speakers processed scrambled sentences more slowly than canonical sentences in the new-given condition, while such processing difficulty associated with scrambled sentences was reduced in the given-new condition. This suggests that information structures can facilitate L1 processing of scrambling, consistent with findings from previous studies (e.g., Jackson, 2008; Park, 2014).

For L2 speakers, although they also showed the given-before-new preferences in the offline task, they did not show an interaction between information structure and word order. In other words, they preferred given-new sentences over new-given sentences regardless of word order, and canonical over scrambled sentences regardless of information structures. In terms of online processing, their adherence to the given-before-new principle was not evident. Across the critical regions, L2 learners tended to prefer the new-given sentences over the given-new sentences, which was interpreted as lexical priming effects in the previous chapter. Although the L2 learners have knowledge of the given-before-new principle as shown in the offline task, they failed to utilize such information during online sentence processing. However, it should be also noted that the L2 speakers demonstrated native-like processing patterns at the gap position (i.e., at the adverb between the second NP and the embedded verb), with longer reading times for scrambled sentences than canonical sentences. This indicates that at least they could process local scrambling by using case-marking information in a native-like manner, while their use of information structure during online processing is different from that of native speakers.

Thus, data from Experiment 3 reveal that even those L2 learners who can utilize morphosyntactic information in processing of local scrambling in a native-like manner have difficulties in integrating syntactic parsing with information structures. This is consistent with what the IH predicts: the locus of L2 processing difficulties is at the interface between syntax and pragmatics.

In terms of learners' proficiency, however, the PCA can raise a question against the IH, asking whether the learners in Experiment 3 can be considered as near-native speakers with the

L2 end-state grammar or non-near-native learners still in the course of development. If the L2 learners in Experiment 3 are not near-native speakers, the PCA can argue that the learners may not attain sufficient proficiency levels to integrate the syntax and pragmatic information during online processing. The proficiency-based account by the PCA can be supported when the L2 learners in Experiments 2 and 3 are compared, as the learners in Experiment 2 are more advanced than those in Experiment 3 in terms of their proficiency test scores ($t(37.093)=2.7844$, $p < 0.05$).

In this regard, it seems difficult to make a strong claim for the IH until testing L2 learners at least at the same proficiency levels as those in Experiment 2 or near-native speakers to dismiss the proficiency issue. Thus, at this point it seems reasonable to conclude that the L2 learners possibly in non-end-state contexts in Experiment 3 demonstrate optionality in integrating knowledge of information structure with scrambling although they can process local scrambling by using morphosyntactic information in a native-like manner.

6.2. Limitations and directions for future study

Before concluding this study, the limitations of the study should be acknowledged. First, due to the complexity of LD scrambling, the online comprehension accuracy rates for L2 learners are quite low in Experiment 2 although their offline comprehension accuracy (the overall accuracy rate was 88%) indicates that they are capable of interpreting LD scrambling. Thus, there is a discrepancy between analyses on all trials and on correctly answered trials. In the discussion section above, results from all trials are considered since our main research interest is in how L2 learners generally process sentences, not only the sentences that are correctly answered in the comprehension task. If the accuracy effect is considered, however, data in this study should be interpreted with caution.

In addition, as discussed above, this study cannot tease apart the two approaches, the IH and the PCA, in Experiment 3. Future study should test L2 learners at about the same proficiency levels as those in Experiment 2 to test the interaction effect between information structure and word order. By doing so, we could provide a strong evidence for either the IH or the PCA. Furthermore, in future research, it would be interesting to test LD scrambling instead of local scrambling in Experiment 3, to examine the effect of working memory on the interaction between syntactic processes associated with the filler-gap dependency and information structures.

In future research, different methodologies can be utilized instead of a self-paced reading task. For example, an eye-tracking method can be used to examine a tendency to predict gaps in scrambled sentences and other linguistic materials (including information structure) as a sentence unfolds. Since such predictive processing with regard to plausibility was observed among native speakers in Experiment 2, using more sensitive methodologies to examine whether L2 learners can also predict upcoming information as native speakers do in processing scrambled sentences, and how working memories and other factors influence such L2 predictive processing, and whether L1 and L2 predictive processing is similar or different.

6.3. Conclusion

The present study examined whether local and LD scrambling affects L1 sentence processing and whether L2 learners of Korean rely on morphosyntactic (case-marking) or non-syntactic (plausibility or heuristic word order) information in processing of scrambling (i.e., the filler-gap dependencies), compared to native Korean speakers. In addition, this study also examined how the knowledge of information structure can interact with syntax in processing of scrambled sentences. In terms of L2 processing, the study contributes to our understanding of whether individual factors such as proficiency and working memory influence native-like processing of scrambling as well as native-like use of morphosyntactic information during online processing.

Data from this study reveal that L2 individual differences in working memory capacities and proficiency influence how they process scrambled sentences. However, it also suggests that the difficulty of integrating syntactic parsing with information structure cannot be overcome in the L2 non-end-state grammar. Overall, the results of the present study are more consistent with what the processing capacity approach (PCA) predicts, L1 and L2 differences can be attributed to processing difficulties rather than learners' incomplete grammatical representations, and thus support the view that L2 processing is quantitatively rather than qualitatively different from L1 processing.

References

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*(3), 191–238.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247−264.

Aoshima, S., Yoshida, M., & Phillips, C. (2009). Incremental processing of coreference and binding in Japanese. *Syntax*, *12*(2), 93–134.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255–278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Beilock, S. L., & DeCaro, M. S. (2007). From poor performance to success under stress: working memory, strategy selection, and mathematical problem solving under pressure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(6), 983–998.

Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language*, *34*(6), 774–806.

Bošković, Z., & Takahashi, D. (1998). Scrambling and last resort. *Linguistic Inquiry, 29*(2), 347–366.

Christianson, K., Hollingworth, A. Halliwell, J. and Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*, 368–407.

Christianson, K., Luke, S., & Ferreira, F. (2010). Effects of Plausibility on Structural Priming. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *36*(2), 537–544.

Christianson, K. (2016). When language comprehension goes wrong for the right reasons: Good-enough, underspecified, or shallow language processing. *Quarterly journal of experimental psychology*, *69*(5), 817–828.

Chung, E. S. (2013). *Exploring the degree of native-likeness in bilingual acquisition: second and heritage language acquisition of Korean case-ellipsis*. Doctoral dissertation, University of Illinois at Urbana-Champaign.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*, 3−42.

Cunnings, I. (2017). Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition*, *20*(4), 659−678.

Dallas, A., DeDe, G., & Nicol, J. (2013). An Event-Related Potential (ERP) Investigation of Filler-Gap Processing in Native and Second Language Speakers. *Language Learning*, *63*(4), 766–799.

Dussias, P. (2001). Sentence parsing in fluent Spanish-English bilinguals. In J. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (pp. 159−176). Malden, MA: Blackwell.

Dussias, P. E., & Piñar, P. (2010). Effects of reading span and plausibility in the reanalysis of wh-gaps by Chinese-English second language speakers. *Second Language Research*, *26*(4), 443–472.

Dussias, P. E., & Scaltz, T. R. C. (2008). Spanish–English L2 speakers' use of subcategorization bias information in the resolution of temporary ambiguity during second language reading. *Acta psychologica*, *128*(3), 501–513.

Erdocia, K., Laka, I., Mestres-Missé, A., & Rodriguez-Fornells, A. (2009). Syntactic complexity and ambiguity resolution in a free word order language: Behavioral and electrophysiological evidences from Basque. *Brain and language*, *109*(1), 1–17.

Erdocia, K., Zawiszewski, A., & Laka, I. (2014). Word order processing in a second language: from VO to OV. *Journal of psycholinguistic research*, *43*(6), 815–837.

Fanselow, G. (1990). Scrambling as NP movement. In G. Grewendorf & W. Sternefeld (Eds.), *Scrambling and barriers* (pp. 113–140). Amsterdam: John Benjamins.

Felser, C., Roberts, L., Gross, R., & Marinis, T. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, *24*, 453–489.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*, 164–203.

Ferreira, F., & Patson, N. D. (2007). The 'good-enough' approach to language comprehension. *Language and Linguistics Compass*, *1*, 71–83.

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PloS one*, *8*(10), e77661.

Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & cognition*, *43*(2), 226–236.

Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of statistical software*, *8*(15), 1–27.

Frazier, L. (1989). Against lexical generation of syntax. In W. D. Marslen-Wilson (Ed.), *Lexical* *representation and process* (pp. 505–528). Cambridge, MA: MIT Press.

Frenck-Mestre, C., & Pynte, J. (1997). Syntactic ambiguity resolution while reading in second and native languages. *The Quarterly Journal of Experimental Psychology*, *50A*, 119–148.

Frenck-Mestre, C. (2002). An on-line look at sentence processing in the second language. In R. Heredia & J. Altarriba (Eds.), *Bilingual sentence processing* (pp. 217–236). New York: Elsevier.

Hagiwara, H., Soshi, T., Ishihara, M., & Imanaka, K. (2007). A topographical study on the event-related potential correlates of scrambled word order in Japanese complex sentences. *Journal of Cognitive Neuroscience*, *19*(2), 175–193.

Havik, E., Roberts, L., Van Hout, R., Schreuder, R., & Haverkort, M. (2009). Processing subject-object ambiguities in the L2: A self-paced reading study with German L2 learners of Dutch. *Language Learning*, *59*(1), 73–112.

Hoji, H., (1985). *Logical Form Constraints and configurational structures in Japanese*. Doctoral dissertation, University of Washington.

Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research*, *22*(3), 369−397.

Hopp, H. (2007). *Ultimate attainment at the interfaces in second language acquisition: Grammar and processing*. Doctoral dissertation, University of Groningen.

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, *120*(4), 901−931.

Hyönä, J. & Hujanen, H. (1997). Effects of Case Marking and Word Order on Sentence Parsing in Finnish: An Eye Fixation Analysis. *The Quarterly Journal of Experimental*

Psychology: A Human Experimental Psychology, 50A(4), 841–858.

Jackson, K. H. (2008). *The effect of information structure on Korean scrambling*. Doctoral dissertation, University of Hawai'i.

Jackson, C. (2008). Proficiency level and the interaction of lexical and morphosyntactic information during L2 sentence processing. *Language Learning, 58*(4), 875–909.

Jackson, C. N., & Roberts, L. (2010). Animacy affects the processing of subject–object ambiguities in the second language: Evidence from self-paced reading with German second language learners of Dutch. *Applied Psycholinguistics, 31*(4), 671–691.

Jackson, C. N., & Van Hell, J. G. (2011). The effects of L2 proficiency level on the processing of wh-questions among Dutch second language speakers of English. *IRAL-International Review of Applied Linguistics in Language Teaching, 49*(3), 195–219.

Jiang, N., (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics, 25*, 603–634.

Juffs, A. (2004). Representation, processing and working memory in a second language. *Transactions of the Philological Society, 102*, 199−225.

Juffs, A. (2005). The influence of first language on the processing of wh-movement in English as a second language. *Second Language Research, 21*(2), 121–151.

Juffs, A., & Harrington, M. (1995). Parsing effects in second language sentence processing: Subject and object asymmetries in wh-extraction. *Studies in second language acquisition, 17*(4), 483–516.

Just, M. A., Carpenter, P., & Woolley, J. D. (1982). Paradigms and processes and in reading comprehension. *Journal of Experimental Psychology: General, 111*(2), 228–238.

Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different?. *Linguistic*

Approaches to Bilingualism, *4*(2), 257–282.

Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, *94*(2), 113–147.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye-movements. *Journal of Memory and Language*, *49*, 133−159.

Kamide, Y., & Mitchell, D. C. (1999). Incremental pre-head attachment in Japanese parsing. *Language and cognitive processes*, *14*(5-6), 631–662.

Kim, Y. (1999). The effects of case marking information on Korean sentence processing. *Language and cognitive processes*, *14*(5-6), 687–714.

Kim, J. H., & Christianson, K. (2017). Working memory effects on L1 and L2 processing of ambiguous relative clauses by Korean L2 learners of English. *Second Language Research*, *33*(3), 365–388.

Koizumi, M., & Imamura, S. (2017). Interaction between syntactic structure and information structure in the processing of a head-final language. *Journal of psycholinguistic research*, *46*(1), 247–260.

Lenth, R. (2018). Emmeans: Estimated marginal means, aka least-squares means. *R Package Version*, *1*(2).

Lim, J. H., & Christianson, K. (2015). Second language sensitivity to agreement errors: Evidence from eye movements during comprehension and translation. *Applied Psycholinguistics*, *36*(6), 1283–1315.

Lim, J-H., & Christianson, K. (2013a). Integrating meaning and structure in L1-L2 and L2-L1 translations. *Second Language Research*, *29*, 233–256.

Lim, J-H., & Christianson, K. (2013b). Second language sentence processing in reading for comprehension and translation. *Bilingualism: Language and Cognition*, *16*(3), 518–537.

Lim, J-H., & Christianson, K. (2015). L2 sensitivity to agreement errors: Evidence from eye movements during comprehension and translation. *Applied Psycholinguistics*, *36*, 1283–1315.

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of verbal learning and verbal behavior*, *23*(2), 127–150.

Mahajan, A. K. (1990). *The A/A-bar distinction and movement theory*. Doctoral dissertation, Massachusetts Institute of Technology.

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition*, *27*, 53−78.

Mazuka, R., Itoh, K., & Kondo, T. (2002). Cost of scrambling in Japanese sentence processing. In M. Nakayama (Ed.), *Sentence processing in East Asian languages* (pp. 131–166). Stanford: CSLI.

McDonald, J. (2006). Beyond the critical period: Processing based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, *55*, 381–401.

Mitsugi, S., & MacWhinney, B. (2010). Second language processing in Japanese scrambled sentences. In B. VanPatten & J. Jegerski (Eds.), *Research in second language processing and parsing* (pp. 159–175). Amsterdam: John Benjamins.

Miyamoto, E. T., & Takahashi, S. (2002). Sources of difficulty in the processing of scrambling in Japanese. In M. Nakayama (Ed.), *Sentence processing in East Asian languages* (pp.167–188). Stanford: CSLI.

Miyamoto, E. T., & Takahashi, S. (2004). Filler-gap dependencies in the processing of scrambling in Japanese. *Language and Linguistics*, *5*, 153–166.

Nakano, Y., Felser, C., & Clahsen, H. (2002). Antecedent priming at trace positions in Japanese long-distance scrambling. *Journal of psycholinguistic research*, *31*(5), 531–571.

Nakayama, M. (1995). Scrambling and probe recognition. In R. Mazuka & N. Nagai (Eds.), *Japanese sentence processing* (pp. 257‒273). Hillsdale, NJ: Lawrence Erlbaum Associates.

Neeleman, A., & Reinhart, T. (1998). Scrambling and the PF interface. In M. Butt & W. Geuder (Eds.), *The projection of arguments: Lexical and compositional factors* (pp. 309–353). Stanford, CA: CSLI.

Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing: A study of relative clause attachment in Greek. *Studies in Second Language Acquisition*, *24*, 501–528.

Park, K. S. (2014). *Information structure and dative word-order alternations in English and Korean: L1 children, L2 children, and L2 adults*. Doctoral dissertation, University of Hawai'i at Mānoa.

Pickering, M. J., & Barry, G. D. (1991). Sentence processing without empty categories. *Language and Cognitive Processes*, *6*, 229‒259.

Pickering, M. J. (1993). Direct association and sentence processing: A reply to Gorrell and to Gibson and Hickok. *Language and Cognitive Processes*, *8*, 163–196.

Pickering, M. J. (1994). Processing local and unbounded dependencies: A unified account. *Journal of Psycholinguistic Research*, 23, 323−352.

Pritchett, B. L. (1991). Head position and parsing ambiguity. *Journal of Psycholinguistic Research*, *20*(3), 251–270.

Roberts, L., & Felser, C. (2011). Plausibility and recovery from garden paths in second language sentence processing. *Applied Psycholinguistics*, *32*(2), 299–331.

Roberts, L., Marinis, T., Felser, C., & Clahsen, H. (2007). Antecedent priming at trace positions in children's sentence processing. *Journal of psycholinguistic research*, *36*(2), 175–188.

Rösler, F., Pechmann, T., Streb, J., Röder, B., & Hennighausen, E. (1998). Parsing of sentences in a language with varying word order: word-by-word variations of processing demands are revealed by even-related potentials. *Journal of Memory and Language*, *38*, 150–176.

Saito, M. (1985). *Some asymmetries in Japanese and their theoretical implications*. Doctoral dissertation, Massachusetts Institute of Technology.

Sekerina, I. (1997). *The syntax and processing of Russian scrambled constructions*. Doctoral Dissertation. City University of New York: NY.

Sekerina, I. A. (2003). Scrambling and processing: dependencies, complexity, and constraints. *Word order and scrambling*, *4*, 301–324.

Shin, S. (2007). Corpus-based study of word order variations in Korean. *Proceedings of the Corpus Linguistics Conference (CL2007)* (pp. 27–30).

Silva, R., Clahsen, H., (2008). Morphologically complex words in L1 and L2 processing: evidence from masked priming in English. *Bilingualism: Language and Cognition*, *11*, 245–260.

Sorace, A., & Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second*

Language Research, *22*(3), 339–368.

Sorace, A. (2011). Pinning down the concept of “interface” in bilingualism. *Linguistic Approaches to Bilingualism*, *1*(1), 1–33.

Sun, C. F., & Givón, T. (1985). On the so-called SOV word order in Mandarin Chinese: A quantified text study and its implications. *Language*, *61*(2), 329–351.

Suzuki, T. (2013). Children’s on-line processing of scrambling in Japanese. *Journal of psycholinguistic research*, *42*(2), 119–137.

Tamaoka, K., Sakai, H., Kawahara, J., & Miyaoka, Y. (2003). The effects of phrase-length order and scrambling in the processing of visually presented Japanese sentences. *Journal of Psycholinguistic Research*, *32*(4), 431–454.

Tamaoka, K., Asano, M., Miyaoka, Y., & Yokosawa, K. (2014). Pre-and post-head processing for single-and double-scrambled sentences of a head-final language as measured by the eye tracking method. *Journal of psycholinguistic research*, *43*(2), 167–185.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(3), 528–553.

Ueno, M., & Kluender, R. (2003). Event-related brain indices of Japanese scrambling. *Brain and Language*, *86*(2), 243−271.

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and cognition*, *4*(02), 105−122.

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times.

Journal of Experimental Psychology: Learning, Memory, and Cognition, *31*, 443−467.

Webelhuth, G. (1989). Syntactic saturation phenomena and the modern Germanic languages. Doctoral dissertation, University of Massachusetts, Amherst.

Williams, J. N. (2006). Incremental interpretation in second language sentence processing. *Bilingualism: Language and Cognition*, *9*(1), 71–88.

Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*. Springer-Verlag, New York.

Yamashita, H. (1997). The effects of word order and case marking information on the processing of Japanese. *Journal of Psycholinguistic Research*, *26*(2), 163–188.

Zhou, P., & Christianson, K. (2016a). I “hear” what you're “saying”: Auditory perceptual simulation, reading speed, and reading comprehension. *The Quarterly Journal of Experimental Psychology*, *69*(5), 972–995.

Zhou, P., & Christianson, K. (2016b). Auditory perceptual simulation: Simulating speech rates or accents?. *Acta psychologica*, *168*, 85–90.

Appendix A

Materials for Experiments 1, 2, and 3

Experiment 1

Target sentences

Condition a: Canonical, Plausible

Condition b: Scrambled, Plausible

Condition c: Canonical, Implausible

Condition d: Scrambled, Implausible

1a. 의사가 환자를 치료했다고 존이 생각했다.

1b. 환자를 의사가 치료했다고 존이 생각했다.

1c. 환자가 의사를 치료했다고 존이 생각했다.

1d. 의사를 환자가 치료했다고 존이 생각했다.

2a. 화가가 대통령을 그렸다고 메리가 생각했다.

2b. 대통령을 화가가 그렸다고 메리가 생각했다.

2c. 대통령이 화가를 그렸다고 메리가 생각했다.

2d. 화가를 대통령이 그렸다고 메리가 생각했다.

3a. 경찰이 도둑을 잡았다고 메리가 생각했다.

3b. 도둑을 경찰이 잡았다고 메리가 생각했다.

3c. 도둑이 경찰을 잡았다고 메리가 생각했다.

3d. 경찰을 도둑이 잡았다고 메리가 생각했다.

4a. 사장이 직원을 해고했다고 리사가 생각했다.

4b. 직원을 사장이 해고했다고 리사가 생각했다.

4c. 직원이 사장을 해고했다고 리사가 생각했다.

4d. 사장을 직원이 해고했다고 리사가 생각했다.

5a. 아빠가 아들을 혼냈다고 존이 생각했다.

5b. 아들을 아빠가 혼냈다고 존이 생각했다.

5c. 아들이 아빠를 혼냈다고 존이 생각했다.

5d. 아빠를 아들이 혼냈다고 존이 생각했다.

6a. 운전사가 손님을 태웠다고 톰이 생각했다.

6b. 손님을 운전사가 태웠다고 톰이 생각했다.

6c. 손님이 운전사를 태웠다고 톰이 생각했다.

6d. 운전사를 손님이 태웠다고 톰이 생각했다.

7a. 스튜어디스가 손님을 안내했다고 메리가 생각했다.

7b. 손님을 스튜어디스가 안내했다고 메리가 생각했다.

7c. 손님이 스튜어디스를 안내했다고 메리가 생각했다.

7d. 스튜어디스를 손님이 안내했다고 메리가 생각했다.

8a. 기자가 군인을 인터뷰했다고 리사가 생각했다.

8b. 군인을 기자가 인터뷰했다고 리사가 생각했다.

8c. 군인이 기자를 인터뷰했다고 리사가 생각했다.

8d. 기자를 군인이 인터뷰했다고 리사가 생각했다.

9a. 사장님이 직원을 혼냈다고 존이 생각했다.

9b.	직원을 사장님이 혼냈다고 존이 생각했다.

9c.	직원이 사장님을 혼냈다고 존이 생각했다.

9d.	사장님을 직원이 혼냈다고 존이 생각했다.

10a.	간호사가 환자를 돌봤다고 톰이 생각했다.

10b.	환자를 간호사가 돌봤다고 톰이 생각했다.

10c.	환자가 간호사를 돌봤다고 톰이 생각했다.

10d.	간호사를 환자가 돌봤다고 톰이 생각했다.

11a.	코치가 선수를 지도했다고 메리가 생각했다.

11b.	선수를 코치가 지도했다고 메리가 생각했다.

11c.	선수가 코치를 지도했다고 메리가 생각했다.

11d.	코치를 선수가 지도했다고 메리가 생각했다.

12a.	사진사가 모델을 찍었다고 리사가 생각했다.

12b.	모델을 사진사가 찍었다고 리사가 생각했다.

12c.	모델이 사진사를 찍었다고 리사가 생각했다.

12d.	사진사를 모델이 찍었다고 리사가 생각했다.

13a.	소방관이 아이를 구했다고 존이 생각했다.

13b.	아이를 소방관이 구했다고 존이 생각했다.

13c.	아이가 소방관을 구했다고 존이 생각했다.

13d.	소방관을 아이가 구했다고 존이 생각했다.

14a.	군인이 노인을 보호했다고 톰이 생각했다.

14b. 노인을 군인이 보호했다고 톰이 생각했다.

14c. 노인이 군인을 보호했다고 톰이 생각했다.

14d. 군인을 노인이 보호했다고 톰이 생각했다.

15a. 선생님이 학생을 가르쳤다고 톰이 생각했다.

15b. 학생을 선생님이 가르쳤다고 톰이 생각했다.

15c. 학생이 선생님을 가르쳤다고 톰이 생각했다.

15d. 선생님을 학생이 가르쳤다고 톰이 생각했다.

16a. 이웃이 도둑을 신고했다고 리사가 생각했다.

16b. 도둑을 이웃이 신고했다고 리사가 생각했다.

16c. 도둑이 이웃을 신고했다고 리사가 생각했다.

16d. 이웃을 도둑이 신고했다고 리사가 생각했다.

17a. 남자가 돼지를 키웠다고 존이 생각했다.

17b. 돼지를 남자가 키웠다고 존이 생각했다.

17c. 돼지가 남자를 키웠다고 존이 생각했다.

17d. 남자를 돼지가 키웠다고 존이 생각했다.

18a. 여자가 고양이를 쓰다듬었다고 톰이 생각했다.

18b. 고양이를 여자가 쓰다듬었다고 톰이 생각했다.

18c. 고양이가 여자를 쓰다듬었다고 톰이 생각했다.

18d. 여자를 고양이가 쓰다듬었다고 톰이 생각했다.

19a. 아이가 거북이를 놓아줬다고 리사가 생각했다.

19b. 거북이를 아이가 놓아줬다고 리사가 생각했다.

19c. 거북이가 아이를 놓아줬다고 리사가 생각했다.

19d. 아이를 거북이가 놓아줬다고 리사가 생각했다.

20a. 남자아이가 강아지를 안았다고 존이 생각했다.

20b. 강아지를 남자아이가 안았다고 존이 생각했다.

20c. 강아지가 남자아이를 안았다고 존이 생각했다.

20d. 남자아이를 강아지가 안았다고 존이 생각했다.

21a. 할아버지가 코끼리를 길들였다고 톰이 생각했다.

21b. 코끼리를 할아버지가 길들였다고 톰이 생각했다.

21c. 코끼리가 할아버지를 길들였다고 톰이 생각했다.

21d. 할아버지를 코끼리가 길들였다고 톰이 생각했다.

22a. 아버지가 새를 만졌다고 메리가 생각했다.

22b. 새를 아버지가 만졌다고 메리가 생각했다.

22c. 새가 아버지를 만졌다고 메리가 생각했다.

22d. 아버지를 새가 만졌다고 메리가 생각했다.

23a. 할머니가 소를 돌봤다고 리사가 생각했다.

23b. 소를 할머니가 돌봤다고 리사가 생각했다.

23c. 소가 할머니를 돌봤다고 리사가 생각했다.

23d. 할머니를 소가 돌봤다고 리사가 생각했다.

24a. 화가가 사자를 그렸다고 존이 생각했다.

24b. 사자를 화가가 그렸다고 존이 생각했다.

24c. 사자가 화가를 그렸다고 존이 생각했다.

24d. 화가를 사자가 그렸다고 존이 생각했다.

25a. 여동생이 호랑이를 피했다고 톰이 생각했다.

25b. 호랑이를 여동생이 피했다고 톰이 생각했다.

25c. 호랑이가 여동생을 피했다고 톰이 생각했다.

25d. 여동생을 호랑이가 피했다고 톰이 생각했다.

26a. 어머니가 토끼를 선택했다고 메리가 생각했다.

26b. 토끼를 어머니가 선택했다고 메리가 생각했다.

26c. 토끼가 어머니를 선택했다고 메리가 생각했다.

26d. 어머니를 토끼가 선택했다고 메리가 생각했다.

27a. 여자아이가 나비를 찾았다고 리사가 생각했다.

27b. 나비를 여자아이가 찾았다고 리사가 생각했다.

27c. 나비가 여자아이를 찾았다고 리사가 생각했다.

27d. 여자아이를 나비가 찾았다고 리사가 생각했다.

28a. 남동생이 비둘기를 구했다고 존이 생각했다.

28b. 비둘기를 남동생이 구했다고 존이 생각했다.

28c. 비둘기가 남동생을 구했다고 존이 생각했다.

28d. 남동생을 비둘기가 구했다고 존이 생각했다.

29a. 노인이 원숭이를 길렀다고 리사가 생각했다.

29b.	원숭이를 노인이 길렀다고 리사가 생각했다.

29c.	원숭이가 노인을 길렀다고 리사가 생각했다.

29d.	노인을 원숭이가 길렀다고 리사가 생각했다.

30a.	초등학생이 물고기를 관찰했다고 메리가 생각했다.

30b.	물고기를 초등학생이 관찰했다고 메리가 생각했다.

30c.	물고기가 초등학생을 관찰했다고 메리가 생각했다.

30d.	초등학생을 물고기가 관찰했다고 메리가 생각했다.

31a.	누나가 병아리를 보살폈다고 톰이 생각했다.

31b.	병아리를 누나가 보살폈다고 톰이 생각했다.

31c.	병아리가 누나를 보살폈다고 톰이 생각했다.

31d.	누나를 병아리가 보살폈다고 톰이 생각했다.

32a.	소년이 말을 탔다고 메리가 생각했다.

32b.	말을 소년이 탔다고 메리가 생각했다.

32c.	말이 소년을 탔다고 메리가 생각했다.

32d.	소년을 말이 탔다고 메리가 생각했다.

33a.	늑대가 양을 기다렸다고 리사가 생각했다.

33b.	양을 늑대가 기다렸다고 리사가 생각했다.

33c.	양이 늑대를 기다렸다고 리사가 생각했다.

33d.	늑대를 양이 기다렸다고 리사가 생각했다.

34a.	사자가 토끼를 잡았다고 톰이 생각했다.

34b. 토끼를 사자가 잡았다고 톰이 생각했다.

34c. 토끼가 사자를 잡았다고 톰이 생각했다.

34d. 사자를 토끼가 잡았다고 톰이 생각했다.

35a. 뱀이 쥐를 먹었다고 메리가 생각했다.

35b. 쥐를 뱀이 먹었다고 메리가 생각했다.

35c. 쥐가 뱀을 먹었다고 메리가 생각했다.

35d. 뱀을 쥐가 먹었다고 메리가 생각했다.

36a. 고양이가 새를 붙잡았다고 리사가 생각했다.

36b. 새를 고양이가 붙잡았다고 리사가 생각했다.

36c. 새가 고양이를 붙잡았다고 리사가 생각했다.

36d. 고양이를 새가 붙잡았다고 리사가 생각했다.

37a. 악어가 말을 물었다고 존이 생각했다.

37b. 말을 악어가 물었다고 존이 생각했다.

37c. 말이 악어를 물었다고 존이 생각했다.

37d. 악어를 말이 물었다고 존이 생각했다.

38a. 개가 오리를 공격했다고 톰이 생각했다.

38b. 오리를 개가 공격했다고 톰이 생각했다.

38c. 오리가 개를 공격했다고 톰이 생각했다.

38d. 개를 오리가 공격했다고 톰이 생각했다.

39a. 펭귄이 물고기를 삼켰다고 메리가 생각했다.

39b. 물고기를 펭귄이 삼켰다고 메리가 생각했다.

39c. 물고기가 펭귄을 삼켰다고 메리가 생각했다.

39d. 펭귄을 물고기가 삼켰다고 메리가 생각했다.

40a. 호랑이가 원숭이를 사냥했다고 리사가 생각했다.

40b. 원숭이를 호랑이가 사냥했다고 리사가 생각했다.

40c. 원숭이가 호랑이를 사냥했다고 리사가 생각했다.

40d. 호랑이를 원숭이가 사냥했다고 리사가 생각했다.

41a. 닭이 병아리를 보호했다고 존이 생각했다.

41b. 병아리를 닭이 보호했다고 존이 생각했다.

41c. 병아리가 닭을 보호했다고 존이 생각했다.

41d. 닭을 병아리가 보호했다고 존이 생각했다.

42a. 사자가 소를 따라갔다고 톰이 생각했다.

42b. 소를 사자가 따라갔다고 톰이 생각했다.

42c. 소가 사자를 따라갔다고 톰이 생각했다.

42d. 사자를 소가 따라갔다고 톰이 생각했다.

43a. 고래가 물고기를 놓쳤다고 메리가 생각했다.

43b. 물고기를 고래가 놓쳤다고 메리가 생각했다.

43c. 물고기가 고래를 놓쳤다고 메리가 생각했다.

43d. 고래를 물고기가 놓쳤다고 메리가 생각했다.

44a. 원숭이가 벌레를 찾았다고 존이 생각했다.

44b.	벌레를 원숭이가 찾았다고 존이 생각했다.

44c.	벌레가 원숭이를 찾았다고 존이 생각했다.

44d.	원숭이를 벌레가 찾았다고 존이 생각했다.

45a.	돼지가 닭을 밀어냈다고 톰이 생각했다.

45b.	닭을 돼지가 밀어냈다고 톰이 생각했다.

45c.	닭이 돼지를 밀어냈다고 톰이 생각했다.

45d.	돼지를 닭이 밀어냈다고 톰이 생각했다.

46a.	소가 오리를 찼다고 메리가 생각했다.

46b.	오리를 소가 찼다고 메리가 생각했다.

46c.	오리가 소를 찼다고 메리가 생각했다.

46d.	소를 오리가 찼다고 메리가 생각했다.

47a.	개가 양을 지켰다고 리사가 생각했다.

47b.	양을 개가 지켰다고 리사가 생각했다.

47c.	양이 개를 지켰다고 리사가 생각했다.

47d.	개를 양이 지켰다고 리사가 생각했다.

48a.	고양이가 쥐를 쫓았다고 존이 생각했다.

48b.	쥐를 고양이가 쫓았다고 존이 생각했다.

48c.	쥐가 고양이를 쫓았다고 존이 생각했다.

48d.	고양이를 쥐가 쫓았다고 존이 생각했다.

Fillers

1. 맛있는 빵을 좋아하는 남자가 매일 빵을 산다.
2. 예쁜 꽃을 좋아하는 여자가 매일 꽃집에 간다.
3. 공부를 열심히 하는 학생이 매일 도서관에 간다.
4. 따뜻한 커피를 좋아하는 여자가 매일 커피를 마신다.
5. 운동을 좋아하는 남자가 친구들과 매일 농구를 한다.
6. 음악을 좋아하는 여자가 하루 만에 바이올린을 배웠다.
7. 공부를 너무 하고 싶었던 할머니가 대학에 입학했다.
8. 과학을 좋아하는 어린아이가 컴퓨터를 한시간 만에 만들었다.
9. 고기를 좋아하는 남자가 50인분을 혼자 모두 먹었다.
10. 눈을 좋아하는 남자아이가 100개의 눈사람을 혼자 만들었다.
11. 남자를 보는 여자가 의자에 앉아 있다.
12. 여자를 좋아하는 남자가 빨간 꽃을 선물했다.
13. 남자를 싫어하는 여자가 전화를 받지 않았다.
14. 돼지를 보는 개가 문 옆에 있다.
15. 개를 좋아하는 돼지가 계속 개를 쳐다본다.
16. 미래를 보는 아이가 내일 일을 이야기한다.
17. 물을 싫어하는 남자가 평생 씻지 않았다.
18. 눈을 좋아하는 여자가 눈으로 옷을 만들었다.
19. 개를 훈련시킨 남자가 개를 매우 무서워했다.

20. 물건을 훔친 도둑이 주인에게 다시 돌려주었다.

21. 아빠가 선물로 준 컴퓨터를 존이 친구에게 팔았다.

22. 엄마가 만들어 준 케이크를 메리가 모두 먹었다.

23. 동생이 그려 준 그림을 존이 벽에 걸었다.

24. 친구가 보낸 이메일을 메리가 실수로 모두 지웠다.

25. 좋아하는 가수가 부른 노래를 존이 따라 불렀다.

26. 할머니가 만들어 주신 목도리를 메리가 여름에만 사용했다.

27. 강아지가 하는 말을 어린아이가 배워서 서로 대화했다.

28. 유명한 과학자가 풀지 못한 문제를 초등학생이 해결했다.

29. 가난한 남자가 주운 1억을 아내가 은행에 저금했다.

30. 아이가 손으로 자른 나무를 아버지가 열심히 모았다.

31. 남자가 싫어하는 여자를 친구들이 파티에 초대했다.

32. 개가 좋아하는 돼지를 주인이 어제 팔았다.

33. 남자가 보고 있는 여자를 사람들이 칭찬했다.

34. 여자가 좋아하는 남자를 친구들이 매우 싫어했다.

35. 돼지가 보고 있는 개를 어린아이가 안았다.

36. 엄마가 혼낸 아이를 아빠가 매우 칭찬했다.

37. 남자가 잡은 도둑을 경찰이 풀어 주었다.

38. 선생님이 칭찬한 아이를 부모가 크게 혼냈다.

39. 기자가 쓴 기사를 아나운서가 마음대로 바꿨다.

40. 가수가 부른 노래를 팬들이 제일 싫어했다.

41. 일요일에 형이 엄마를 도왔다고 존이 말했다.

42. 아침에 언니가 아빠와 통화했다고 메리가 말했다.

43. 밤에 남자가 친구를 만났다고 톰이 말했다.

44. 식당에서 남자가 피자를 시켰다고 존이 말했다.

45. 아침에 누나가 선물을 보냈다고 존이 말했다.

46. 어제 형이 편지를 받았다고 톰이 말했다.

47. 7월에 여자가 우주로 떠났다고 리사가 말했다.

48. 교실에서 초등학생들이 요리를 했다고 리사가 말했다.

49. 거실에서 남자가 강아지와 공부했다고 메리가 말했다.

50. 여름에 여자가 코트를 입었다고 존이 말했다.

51. 오늘 어린아이가 자동차를 샀다고 톰이 말했다.

52. 거실에서 남자가 이불을 빨았다고 리사가 말했다.

53. 공원에서 할아버지가 할머니와 다정하게 산책했다고 메리가 말했다.

54. 방안에서 아버지가 어머니와 즐겁게 이야기했다고 존이 말했다.

55. 길거리에서 남자가 시끄럽게 여자와 싸웠다고 톰이 말했다.

56. 오늘 친구가 어려운 시험을 본다고 리사가 말했다.

57. 교실에서 학생이 더러운 책상을 닦았다고 메리가 말했다.

58. 어제 요리사가 맛있는 요리를 만들었다고 메리가 말했다.

59. 극장에서 여자가 슬픈 영화를 만들었다고 톰이 말했다.

60. 도서관에서 동생이 재밌는 책을 찢었다고 리사가 말했다.

61. 겨울에 남자가 수영으로 하와이에 갔다고 메리가 말했다.

62. 방안에서 여자가 즐겁게 고양이와 숙제했다고 존이 말했다.

63. 바다에서 친구가 열심히 강아지와 수영했다고 톰이 말했다.

64. 산에서 언니가 다정하게 새와 이야기했다고 리사가 말했다.

65. 존이 청소를 하고 메리가 빨래를 했다.

66. 늦잠을 잤지만 메리는 학교에 늦지 않았다.

67. 리사는 도서관에 갔지만 존은 집에서 놀았다.

68. 리사가 피자는 먹었지만 스파게티는 안 먹었다.

69. 메리는 아침 운동을 하고 학교에 간다.

70. 고등학생이 6개월 동안 공부해서 대학에 갔다.

71. 남자가 10년 동안 회사에 지각하지 않았다.

72. 학생이 12년 동안 학교에 결석하지 않았다.

73. 주부가 하루도 빠짐없이 집안을 깨끗하게 청소했다.

74. 여자가 가족들의 옷을 전부 만들어 주었다.

75. 대학생이 컴퓨터로 이메일을 보낼 줄 몰랐다.

76. 감기에 걸려서 아팠지만 메리는 열심히 시험공부를 했다.

77. 눈이 와서 미끄러운 길을 존이 천천히 운전했다.

78. 날씨가 너무 추워서 리사는 두꺼운 코트를 입었다.

79. 돈이 없었지만 메리는 친구의 생일 선물을 샀다.

80. 리사는 너무 피곤해서 수업에 집중할 수 없었다.

81. 톰이 지갑을 잃어버려서 메리에게 돈을 조금 빌렸다.

82. 부자인 사장이 전재산을 모두 학교에 몰래 기부했다.

83. 사랑에 빠진 여자가 1주일 만에 결혼을 결심했다.

84. 10살 때 만난 첫사랑과 남자가 결혼을 했다.

85. 여자가 10년 동안 매일매일 등산을 해서 건강해졌다.

86. 할머니가 멀미가 심해서 차를 탈 수 없었다.

87. 어제 메리와 톰이 거실에서 두꺼운 책을 읽었다.

88. 친구의 생일날 존과 리사가 선물을 샀다.

89. 크리스마스에 톰과 존이 각각 카드를 썼다.

90. 수업 시간에 메리와 리사가 함께 울었다.

91. 산에 올라가서 리사와 톰이 나무를 모두 뽑았다.

92. 존이 점심으로 작은 케이크와 쿠키를 맛있게 먹었다.

93. 리사가 큰 정원에 나무와 꽃을 많이 심었다.

94. 메리가 매일 새벽에 라면과 밥을 함께 먹었다.

95. 톰이 도서관에서 책과 잡지를 사람들에게 팔았다.

96. 존이 영화와 드라마를 1년 동안 보았다.

Experiment 2

Target sentences

Condition a: Canonical, Plausible

Condition b: Scrambled, Plausible

Condition c: Canonical, Implausible

Condition d: Scrambled, Implausible

1a. 나는 어제 의사가 병원에서 환자를 열심히 치료했다고 혼자 생각했다.

1b. 환자를 나는 어제 의사가 병원에서 열심히 치료했다고 혼자 생각했다.

1c. 나는 어제 환자가 병원에서 의사를 열심히 치료했다고 혼자 생각했다.

1d. 의사를 나는 어제 환자가 병원에서 열심히 치료했다고 혼자 생각했다.

2a. 나는 오후에 화가가 집에서 대통령을 대충 그렸다고 솔직히 생각했다.

2b. 대통령을 나는 오후에 화가가 집에서 대충 그렸다고 솔직히 생각했다.

2c. 나는 오후에 대통령이 집에서 화가를 대충 그렸다고 솔직히 생각했다.

2d. 화가를 나는 오후에 대통령이 집에서 대충 그렸다고 솔직히 생각했다.

3a. 나는 밤에 경찰이 골목에서 도둑을 겨우 잡았다고 잠시 생각했다.

3b. 도둑을 나는 밤에 경찰이 골목에서 겨우 잡았다고 잠시 생각했다.

3c. 나는 밤에 도둑이 골목에서 경찰을 겨우 잡았다고 잠시 생각했다.

3d. 경찰을 나는 밤에 도둑이 골목에서 겨우 잡았다고 잠시 생각했다.

4a. 나는 작년에 사장이 회사에서 직원을 냉정하게 해고했다고 줄곧 생각했다.

4b. 직원을 나는 작년에 사장이 회사에서 냉정하게 해고했다고 줄곧 생각했다.

4c. 나는 작년에 직원이 회사에서 사장을 냉정하게 해고했다고 줄곧 생각했다.

4d. 사장을 나는 작년에 직원이 회사에서 냉정하게 해고했다고 줄곧 생각했다.

5a. 나는 아침에 아빠가 큰소리로 아들을 심하게 혼냈다고 멋대로 생각했다.

5b. 아들을 나는 아침에 아빠가 큰소리로 심하게 혼냈다고 멋대로 생각했다.

5c. 나는 아침에 아들이 큰소리로 아빠를 심하게 혼냈다고 멋대로 생각했다.

5d. 아빠를 나는 아침에 아들이 큰소리로 심하게 혼냈다고 멋대로 생각했다.

6a. 나는 일요일에 운전사가 도로에서 손님을 강제로 태웠다고 여전히 생각했다.

6b. 손님을 나는 일요일에 운전사가 도로에서 강제로 태웠다고 여전히 생각했다.

6c. 나는 일요일에 손님이 도로에서 운전사를 강제로 태웠다고 여전히 생각했다.

6d. 운전사를 나는 일요일에 손님이 도로에서 강제로 태웠다고 여전히 생각했다.

7a. 나는 방금 스튜어디스가 비행기에서 손님을 친절하게 안내했다고 잠깐 생각했다.

7b. 손님을 나는 방금 스튜어디스가 비행기에서 친절하게 안내했다고 잠깐 생각했다.

7c. 나는 방금 손님이 비행기에서 스튜어디스를 친절하게 안내했다고 잠깐 생각했다.

7d. 스튜어디스를 나는 방금 손님이 비행기에서 친절하게 안내했다고 잠깐 생각했다.

8a. 나는 저녁에 기자가 사무실에서 군인을 짧게 인터뷰했다고 단순히 생각했다.

8b. 군인을 나는 저녁에 기자가 사무실에서 짧게 인터뷰했다고 단순히 생각했다.

8c. 나는 저녁에 군인이 사무실에서 기자를 짧게 인터뷰했다고 단순히 생각했다.

8d. 기자를 나는 저녁에 군인이 사무실에서 짧게 인터뷰했다고 단순히 생각했다.

9a. 나는 아까 회장님이 회의실에서 사원을 크게 혼냈다고 심각하게 생각했다.

9b. 사원을 나는 아까 회장님이 회의실에서 크게 혼냈다고 심각하게 생각했다.

9c. 나는 아까 사원이 회의실에서 회장님을 크게 혼냈다고 심각하게 생각했다.

9d. 회장님을 나는 아까 사원이 회의실에서 크게 혼냈다고 심각하게 생각했다.

10a. 나는 오늘 간호사가 병실에서 환자를 따뜻하게 돌봤다고 맘속으로 생각했다.

10b. 환자를 나는 오늘 간호사가 병실에서 따뜻하게 돌봤다고 맘속으로 생각했다.

10c. 나는 오늘 환자가 병실에서 간호사를 따뜻하게 돌봤다고 맘속으로 생각했다.

10d. 간호사를 나는 오늘 환자가 병실에서 따뜻하게 돌봤다고 맘속으로 생각했다.

11a. 나는 휴일에 코치가 운동장에서 선수를 성실하게 지도했다고 가끔 생각했다.

11b. 선수를 나는 휴일에 코치가 운동장에서 성실하게 지도했다고 가끔 생각했다.

11c. 나는 휴일에 선수가 운동장에서 코치를 성실하게 지도했다고 가끔 생각했다.

11d. 코치를 나는 휴일에 선수가 운동장에서 성실하게 지도했다고 가끔 생각했다.

12a. 나는 지난주에 사진사가 바닷가에서 모델을 멋지게 찍었다고 때때로 생각했다.

12b. 모델을 나는 지난주에 사진사가 바닷가에서 멋지게 찍었다고 때때로 생각했다.

12c. 나는 지난주에 모델이 바닷가에서 사진사를 멋지게 찍었다고 때때로 생각했다.

12d. 사진사를 나는 지난주에 모델이 바닷가에서 멋지게 찍었다고 때때로 생각했다.

13a. 나는 지난달에 소방관이 산에서 아이를 신속하게 구했다고 항상 생각했다.

13b. 아이를 나는 지난달에 소방관이 산에서 신속하게 구했다고 항상 생각했다.

13c. 나는 지난달에 아이가 산에서 소방관을 신속하게 구했다고 항상 생각했다.

13d. 소방관을 나는 지난달에 아이가 산에서 신속하게 구했다고 항상 생각했다.

14a. 나는 겨울에 군인이 거리에서 노인을 제대로 보호했다고 자주 생각했다.

14b. 노인을 나는 겨울에 군인이 거리에서 제대로 보호했다고 자주 생각했다.

14c. 나는 겨울에 노인이 거리에서 군인을 제대로 보호했다고 자주 생각했다.

14d. 군인을 나는 겨울에 노인이 거리에서 제대로 보호했다고 자주 생각했다.

15a. 나는 오전에 선생님이 교실에서 학생을 열정적으로 가르쳤다고 또다시 생각했다.

15b. 학생을 나는 오전에 선생님이 교실에서 열정적으로 가르쳤다고 또다시 생각했다.

15c. 나는 오전에 학생이 교실에서 선생님을 열정적으로 가르쳤다고 또다시 생각했다.

15d. 선생님을 나는 오전에 학생이 교실에서 열정적으로 가르쳤다고 또다시 생각했다.

16a. 나는 지난번에 이웃이 전화로 도둑을 급하게 신고했다고 간단히 생각했다.

16b. 도둑을 나는 지난번에 이웃이 전화로 급하게 신고했다고 간단히 생각했다.

16c. 나는 지난번에 도둑이 전화로 이웃을 급하게 신고했다고 간단히 생각했다.

16d. 이웃을 나는 지난번에 도둑이 전화로 급하게 신고했다고 간단히 생각했다.

17a. 나는 여름에 남자가 마당에서 돼지를 대충 키웠다고 두고두고 생각했다.

17b. 돼지를 나는 여름에 남자가 마당에서 대충 키웠다고 두고두고 생각했다.

17c. 나는 여름에 돼지가 마당에서 남자를 대충 키웠다고 두고두고 생각했다.

17d. 남자를 나는 여름에 돼지가 마당에서 대충 키웠다고 두고두고 생각했다.

18a. 나는 낮에 여자가 밖에서 고양이를 조심스럽게 쓰다듬었다고 가볍게 생각했다.

18b. 고양이를 나는 낮에 여자가 밖에서 조심스럽게 쓰다듬었다고 가볍게 생각했다.

18c. 나는 낮에 고양이가 밖에서 여자를 조심스럽게 쓰다듬었다고 가볍게 생각했다.

18d. 여자를 나는 낮에 고양이가 밖에서 조심스럽게 쓰다듬었다고 가볍게 생각했다.

19a. 나는 봄에 아이가 연못에 거북이를 그냥 놓아줬다고 자꾸만 생각했다.

19b. 거북이를 나는 봄에 아이가 연못에 그냥 놓아줬다고 자꾸만 생각했다.

19c. 나는 봄에 거북이가 연못에 아이를 그냥 놓아줬다고 자꾸만 생각했다.

19d. 아이를 나는 봄에 거북이가 연못에 그냥 놓아줬다고 자꾸만 생각했다.

20a. 나는 주말에 남자아이가 놀이터에서 강아지를 세게 안았다고 거듭 생각했다.

20b. 강아지를 나는 주말에 남자아이가 놀이터에서 세게 안았다고 거듭 생각했다.

20c. 나는 주말에 강아지가 놀이터에서 남자아이를 세게 안았다고 거듭 생각했다.

20d. 남자아이를 나는 주말에 강아지가 놀이터에서 세게 안았다고 거듭 생각했다.

21a. 나는 매일 할아버지가 동물원에서 코끼리를 힘들게 길들였다고 변함없이 생각했다.

21b. 코끼리를 나는 매일 할아버지가 동물원에서 힘들게 길들였다고 변함없이 생각했다.

21c. 나는 매일 코끼리가 동물원에서 할아버지를 힘들게 길들였다고 변함없이 생각했다.

21d. 할아버지를 나는 매일 코끼리가 동물원에서 힘들게 길들였다고 변함없이

생각했다.

22a. 나는 가을에 아버지가 공원에서 새를 거칠게 만졌다고 조용히 생각했다.

22b. 새를 나는 가을에 아버지가 공원에서 거칠게 만졌다고 조용히 생각했다.

22c. 나는 가을에 새가 공원에서 아버지를 거칠게 만졌다고 조용히 생각했다.

22d. 아버지를 나는 가을에 새가 공원에서 거칠게 만졌다고 조용히 생각했다.

23a. 나는 옛날에 할머니가 창고에서 소를 정성스럽게 돌봤다고 말없이 생각했다.

23b. 소를 나는 옛날에 할머니가 창고에서 정성스럽게 돌봤다고 말없이 생각했다.

23c. 나는 옛날에 소가 창고에서 할머니를 정성스럽게 돌봤다고 말없이 생각했다.

23d. 할머니를 나는 옛날에 소가 창고에서 정성스럽게 돌봤다고 말없이 생각했다.

24a. 나는 예전에 화가가 산속에서 사자를 몰래 그렸다고 종종 생각했다.

24b. 사자를 나는 예전에 화가가 산속에서 몰래 그렸다고 종종 생각했다.

24c. 나는 예전에 사자가 산속에서 화가를 몰래 그렸다고 종종 생각했다.

24d. 화가를 나는 예전에 사자가 산속에서 몰래 그렸다고 종종 생각했다.

25a. 나는 과거에 여동생이 용감하게 호랑이를 재빨리 피했다고 긍정적으로 생각했다.

25b. 호랑이를 나는 과거에 여동생이 용감하게 재빨리 피했다고 긍정적으로 생각했다.

25c. 나는 과거에 호랑이가 용감하게 여동생을 재빨리 피했다고 긍정적으로 생각했다.

25d. 여동생을 나는 과거에 호랑이가 용감하게 재빨리 피했다고 긍정적으로

생각했다.

26a. 나는 월요일에 어머니가 먼저 토끼를 적극적으로 선택했다고 속으로 생각했다.

26b. 토끼를 나는 월요일에 어머니가 먼저 적극적으로 선택했다고 속으로 생각했다.

26c. 나는 월요일에 토끼가 먼저 어머니를 적극적으로 선택했다고 속으로 생각했다.

26d. 어머니를 나는 월요일에 토끼가 먼저 적극적으로 선택했다고 속으로 생각했다.

27a. 나는 화요일에 여자아이가 꽃밭에서 나비를 어렵게 찾았다고 계속 생각했다.

27b. 나비를 나는 화요일에 여자아이가 꽃밭에서 어렵게 찾았다고 계속 생각했다.

27c. 나는 화요일에 나비가 꽃밭에서 여자아이를 어렵게 찾았다고 계속 생각했다.

27d. 여자아이를 나는 화요일에 나비가 꽃밭에서 어렵게 찾았다고 계속 생각했다.

28a. 나는 수요일에 남동생이 지붕에서 비둘기를 조심스럽게 구했다고 수없이 생각했다.

28b. 비둘기를 나는 수요일에 남동생이 지붕에서 조심스럽게 구했다고 수없이 생각했다.

28c. 나는 수요일에 비둘기가 지붕에서 남동생을 조심스럽게 구했다고 수없이 생각했다.

28d. 남동생을 나는 수요일에 비둘기가 지붕에서 조심스럽게 구했다고 수없이 생각했다.

29a. 나는 전에 노인이 집안에서 원숭이를 정성껏 길렀다고 진심으로 생각했다.

29b. 원숭이를 나는 전에 노인이 집안에서 정성껏 길렀다고 진심으로 생각했다.

29c. 나는 전에 원숭이가 집안에서 노인을 정성껏 길렀다고 진심으로 생각했다.

29d. 노인을 나는 전에 원숭이가 집안에서 정성껏 길렀다고 진심으로 생각했다.

30a. 나는 목요일에 초등학생이 수족관에서 물고기를 자세히 관찰했다고 다시 생각했다.

30b. 물고기를 나는 목요일에 초등학생이 수족관에서 자세히 관찰했다고 다시 생각했다.

30c. 나는 목요일에 물고기가 수족관에서 초등학생을 자세히 관찰했다고 다시 생각했다.

30d. 초등학생을 나는 목요일에 물고기가 수족관에서 자세히 관찰했다고 다시 생각했다.

31a. 나는 오래전에 누나가 베란다에서 병아리를 정성껏 보살폈다고 쉽게 생각했다.

31b. 병아리를 나는 오래전에 누나가 베란다에서 정성껏 보살폈다고 쉽게 생각했다.

31c. 나는 오래전에 병아리가 베란다에서 누나를 정성껏 보살폈다고 쉽게 생각했다.

31d. 누나를 나는 오래전에 병아리가 베란다에서 정성껏 보살폈다고 쉽게 생각했다.

32a. 나는 금요일에 소년이 농장에서 말을 즐겁게 탔다고 자유롭게 생각했다.

32b. 말을 나는 금요일에 소년이 농장에서 즐겁게 탔다고 자유롭게 생각했다.

32c. 나는 금요일에 말이 농장에서 소년을 즐겁게 탔다고 자유롭게 생각했다.

32d. 소년을 나는 금요일에 말이 농장에서 즐겁게 탔다고 자유롭게 생각했다.

33a. 나는 밤에 늑대가 뒤에서 양을 조용히 기다렸다고 잠시 생각했다.

33b. 양을 나는 밤에 늑대가 뒤에서 조용히 기다렸다고 잠시 생각했다.

33c. 나는 밤에 양이 뒤에서 늑대를 조용히 기다렸다고 잠시 생각했다.

33d. 늑대를 나는 밤에 양이 뒤에서 조용히 기다렸다고 잠시 생각했다.

34a. 나는 낮에 사자가 풀숲에서 토끼를 쉽게 잡았다고 차분히 생각했다.

34b. 토끼를 나는 낮에 사자가 풀숲에서 쉽게 잡았다고 차분히 생각했다.

34c. 나는 낮에 토끼가 풀숲에서 사자를 쉽게 잡았다고 차분히 생각했다.

34d. 사자를 나는 낮에 토끼가 풀숲에서 쉽게 잡았다고 차분히 생각했다.

35a. 나는 평소에 뱀이 통째로 쥐를 천천히 먹었다고 냉정하게 생각했다.

35b. 쥐를 나는 평소에 뱀이 통째로 천천히 먹었다고 냉정하게 생각했다.

35c. 나는 평소에 쥐가 통째로 뱀을 천천히 먹었다고 냉정하게 생각했다.

35d. 뱀을 나는 평소에 쥐가 통째로 천천히 먹었다고 냉정하게 생각했다.

36a. 나는 어제 고양이가 길에서 새를 재빨리 붙잡았다고 단순하게 생각했다.

36b. 새를 나는 어제 고양이가 길에서 재빨리 붙잡았다고 단순하게 생각했다.

36c. 나는 어제 새가 길에서 고양이를 재빨리 붙잡았다고 단순하게 생각했다.

36d. 고양이를 나는 어제 새가 길에서 재빨리 붙잡았다고 단순하게 생각했다.

37a. 나는 새벽에 악어가 물가에서 말을 사납게 물었다고 진지하게 생각했다.

37b. 말을 나는 새벽에 악어가 물가에서 사납게 물었다고 진지하게 생각했다.

37c. 나는 새벽에 말이 물가에서 악어를 사납게 물었다고 진지하게 생각했다.

37d. 악어를 나는 새벽에 말이 물가에서 사납게 물었다고 진지하게 생각했다.

38a. 나는 토요일에 개가 사냥에서 오리를 갑자기 공격했다고 사실 생각했다.

38b. 오리를 나는 토요일에 개가 사냥에서 갑자기 공격했다고 사실 생각했다.

38c. 나는 토요일에 오리가 사냥에서 개를 갑자기 공격했다고 사실 생각했다.

38d. 개를 나는 토요일에 오리가 사냥에서 갑자기 공격했다고 사실 생각했다.

39a. 나는 오늘 펭귄이 물속에서 물고기를 한입에 삼켰다고 간단하게 생각했다.

39b. 물고기를 나는 오늘 펭귄이 물속에서 한입에 삼켰다고 간단하게 생각했다.

39c. 나는 오늘 물고기가 물속에서 펭귄을 한입에 삼켰다고 간단하게 생각했다.

39d. 펭귄을 나는 오늘 물고기가 물속에서 한입에 삼켰다고 간단하게 생각했다.

40a. 나는 어젯밤에 호랑이가 들판에서 원숭이를 힘들게 사냥했다고 혼자서 생각했다.

40b. 원숭이를 나는 어젯밤에 호랑이가 들판에서 힘들게 사냥했다고 혼자서 생각했다.

40c. 나는 어젯밤에 원숭이가 들판에서 호랑이를 힘들게 사냥했다고 혼자서 생각했다.

40d. 호랑이를 나는 어젯밤에 원숭이가 들판에서 힘들게 사냥했다고 혼자서 생각했다.

41a. 나는 밤늦게 닭이 혼자서 병아리를 철저하게 보호했다고 깊이 생각했다.

41b. 병아리를 나는 밤늦게 닭이 혼자서 철저하게 보호했다고 깊이 생각했다.

41c. 나는 밤늦게 병아리가 혼자서 닭을 철저하게 보호했다고 깊이 생각했다.

41d. 닭을 나는 밤늦게 병아리가 혼자서 철저하게 보호했다고 깊이 생각했다.

42a. 나는 이미 사자가 물가에서부터 소를 바짝 따라갔다고 마음대로 생각했다.

42b. 소를 나는 이미 사자가 물가에서부터 바짝 따라갔다고 마음대로 생각했다.

42c. 나는 이미 소가 물가에서부터 사자를 바짝 따라갔다고 마음대로 생각했다.

42d. 사자를 나는 이미 소가 물가에서부터 바짝 따라갔다고 마음대로 생각했다.

43a. 나는 방금 고래가 눈앞에서 물고기를 실수로 놓쳤다고 실제로 생각했다.

43b. 물고기를 나는 방금 고래가 눈앞에서 실수로 놓쳤다고 실제로 생각했다.

43c. 나는 방금 물고기가 눈앞에서 고래를 실수로 놓쳤다고 실제로 생각했다.

43d. 고래를 나는 방금 물고기가 눈앞에서 실수로 놓쳤다고 실제로 생각했다.

44a. 나는 한낮에 원숭이가 땅에서 벌레를 간신히 찾았다고 내내 생각했다.

44b. 벌레를 나는 한낮에 원숭이가 땅에서 간신히 찾았다고 내내 생각했다.

44c. 나는 한낮에 벌레가 땅에서 원숭이를 간신히 찾았다고 내내 생각했다.

44d. 원숭이를 나는 한낮에 벌레가 땅에서 간신히 찾았다고 내내 생각했다.

45a. 나는 한밤중에 돼지가 몸으로 닭을 강하게 밀어냈다고 일단 생각했다.

45b. 닭을 나는 한밤중에 돼지가 몸으로 강하게 밀어냈다고 일단 생각했다.

45c. 나는 한밤중에 닭이 몸으로 돼지를 강하게 밀어냈다고 일단 생각했다.

45d. 돼지를 나는 한밤중에 닭이 몸으로 강하게 밀어냈다고 일단 생각했다.

46a. 나는 먼저 소가 앞에서 오리를 힘껏 찼다고 끊임없이 생각했다.

46b. 오리를 나는 먼저 소가 앞에서 힘껏 찼다고 끊임없이 생각했다.

46c. 나는 먼저 오리가 앞에서 소를 힘껏 찼다고 끊임없이 생각했다.

46d. 소를 나는 먼저 오리가 앞에서 힘껏 찼다고 끊임없이 생각했다.

47a. 나는 처음부터 개가 옆에서 양을 꿋꿋이 지켰다고 머릿속으로 생각했다.

47b. 양을 나는 처음부터 개가 옆에서 꿋꿋이 지켰다고 머릿속으로 생각했다.

47c. 나는 처음부터 양이 옆에서 개를 꿋꿋이 지켰다고 머릿속으로 생각했다.

47d.	개를 나는 처음부터 양이 옆에서 꿋꿋이 지켰다고 머릿속으로 생각했다.

48a.	나는 한때 고양이가 복도에서 쥐를 재미로 쫓았다고 기본적으로 생각했다.

48b.	쥐를 나는 한때 고양이가 복도에서 재미로 쫓았다고 기본적으로 생각했다.

48c.	나는 한때 쥐가 복도에서 고양이를 재미로 쫓았다고 기본적으로 생각했다.

48d.	고양이를 나는 한때 쥐가 복도에서 재미로 쫓았다고 기본적으로 생각했다.

Fillers

1. 학생이 12년 동안 매일 학교에 갔다.
2. 주부가 하루도 빠짐없이 집안을 깨끗하게 청소했다.
3. 여자가 가족들의 옷을 전부 만들어 주었다.
4. 대학생이 컴퓨터로 이메일을 보낼 줄 몰랐다.
5. 밤에 남자가 친구를 만났다고 형식이가 말했다.
6. 식당에서 남자가 피자를 시켰다고 경호가 말했다.
7. 아침에 누나가 선물을 보냈다고 민수가 말했다.
8. 어제 형이 편지를 받았다고 수근이가 말했다.
9. 민호가 청소를 하고 수정이가 빨래를 했다.
10. 늦잠을 잤지만 정아는 학교에 늦지 않았다.
11. 영희는 도서관에 갔지만 민호는 집에서 놀았다.
12. 지희가 피자는 먹었지만 스파게티는 안 먹었다.
13. 수희는 아침 운동을 하고 학교에 간다.
14. 친구의 생일날 명수와 은정이가 선물을 샀다.
15. 크리스마스에 경민이와 재석이가 각각 카드를 썼다.
16. 눈을 좋아하는 여자가 눈으로 옷을 만들었다.
17. 개를 훈련시킨 남자가 개를 매우 무서워했다.
18. 물건을 훔친 도둑이 주인에게 다시 돌려주었다.
19. 엄마가 혼낸 아이를 아빠가 매우 칭찬했다.

20. 남자가 잡은 도둑을 경찰이 풀어 주었다.

21. 선생님이 칭찬한 아이를 부모가 크게 혼냈다.

22. 기자가 쓴 기사를 아나운서가 일부러 숨겼다.

23. 가수가 부른 노래를 팬들이 제일 싫어했다.

24. 7월에 여자가 우주로 떠났다고 현정이가 말했다.

25. 교실에서 초등학생들이 요리를 했다고 정아가 말했다.

26. 거실에서 남자가 강아지와 공부했다고 현아가 말했다.

27. 여름에 여자가 코트를 입었다고 희준이가 말했다.

28. 오늘 어린아이가 자동차를 샀다고 혁이가 말했다.

29. 거실에서 남자가 이불을 빨았다고 은이가 말했다.

30. 수업 시간에 수정이와 은희가 함께 울었다.

31. 진호가 도서관에서 책과 잡지를 사람들에게 팔았다.

32. 민철이가 영화와 드라마를 전혀 보지 않았다.

33. 남자를 보는 여자가 의자에 앉아 있다.

34. 여자를 좋아하는 남자가 빨간 꽃을 선물했다.

35. 남자를 싫어하는 여자가 전화를 받지 않았다.

36. 돼지를 보는 개가 문 옆에 있다.

37. 개를 좋아하는 돼지가 계속 개를 쳐다본다.

38. 남자가 보고 있는 여자를 사람들이 칭찬했다.

39. 여자가 좋아하는 남자를 친구들이 매우 싫어했다.

40. 돼지가 보고 있는 개를 어린아이가 안았다.

41. 음악을 좋아하는 여자가 하루 만에 바이올린을 배웠다.

42. 공부를 너무 하고 싶었던 할머니가 대학에 입학했다.

43. 과학을 좋아하는 어린아이가 컴퓨터를 1시간 만에 만들었다.

44. 고기를 좋아하는 남자가 50인분을 혼자 모두 먹었다.

45. 눈을 좋아하는 남자아이가 100개의 눈사람을 혼자 만들었다.

46. 유명한 과학자가 풀지 못한 문제를 초등학생이 해결했다.

47. 가난한 남자가 주운 1억을 아내가 은행에 저금했다.

48. 아이가 손으로 자른 나무를 아버지가 열심히 모았다.

49. 극장에서 여자가 슬픈 영화를 만들었다고 수호가 말했다.

50. 도서관에서 동생이 재밌는 책을 찢었다고 지희가 말했다.

51. 겨울에 남자가 수영으로 하와이에 갔다고 민희가 말했다.

52. 방안에서 여자가 즐겁게 고양이와 숙제했다고 민호가 말했다.

53. 바다에서 친구가 열심히 강아지와 수영했다고 재현이가 말했다.

54. 산에서 언니가 다정하게 새와 이야기했다고 영희가 말했다.

55. 산에 올라가서 미영이와 영호가 나무를 모두 뽑았다.

56. 민정이가 매일 새벽에 라면과 밥을 함께 먹었다.

57. 맛있는 빵을 좋아하는 남자가 매일 빵을 산다.

58. 예쁜 꽃을 좋아하는 여자가 매일 꽃집에 간다.

59. 공부를 열심히 하는 학생이 매일 도서관에 간다.

60. 아빠가 선물로 준 컴퓨터를 철수가 친구에게 팔았다.

61. 엄마가 만들어 준 케이크를 영미가 모두 먹었다.

62. 동생이 그려 준 그림을 민수가 벽에 걸었다.

63. 친구가 보낸 이메일을 수미가 실수로 모두 지웠다.

64. 좋아하는 가수가 부른 노래를 영철이가 따라 불렀다.

65. 어제 민희와 경수가 거실에서 두꺼운 책을 읽었다.

66. 10살 때 만난 첫사랑과 남자가 결혼을 했다.

67. 여자가 10년 동안 매일매일 등산을 해서 건강해졌다.

68. 할머니가 멀미가 심해서 차를 탈 수 없었다.

69. 길거리에서 남자가 시끄럽게 여자와 싸웠다고 희철이가 말했다.

70. 오늘 친구가 어려운 시험을 본다고 주희가 말했다.

71. 교실에서 학생이 더러운 책상을 닦았다고 정아가 말했다.

72. 어제 요리사가 맛있는 요리를 만들었다고 미정이가 말했다.

73. 감기에 걸려서 아팠지만 수미는 열심히 시험공부를 했다.

74. 눈이 와서 미끄러운 길을 민수가 천천히 운전했다.

75. 날씨가 너무 추워서 영미는 두꺼운 코트를 입었다.

76. 돈이 없었지만 다영이는 친구의 생일 선물을 샀다.

77. 준희는 너무 피곤해서 수업에 집중할 수 없었다.

78. 민수가 지갑을 잃어버려서 수희에게 돈을 조금 빌렸다.

79. 준호가 점심으로 작은 케이크와 쿠키를 맛있게 먹었다.

80. 민경이가 큰 정원에 나무와 꽃을 많이 심었다.

81. 할머니가 만들어 주신 두꺼운 목도리를 지혜가 여름에만 사용했다.

82. 강아지가 하는 말을 어린아이가 잘 배워서 서로 대화했다.

83. 고등학생이 6개월 동안 열심히 공부해서 좋은 대학에 갔다.

84. 남자가 일이 좋아서 평생 한 번도 쉬지 않았다.

85. 성훈이에게 친구가 뱀으로 만든 요리를 깜짝 선물로 주었다.

86. 미경이에게 가수가 직접 집으로 찾아와 노래를 불러 주었다.

87. 주호는 한 번 들은 노래도 다 따라 불렀다.

88. 성민이는 길에서 우연히 마주친 사람의 얼굴을 다 기억했다.

89. 텔레비전으로 지혜가 방금 본 뉴스에 지혜의 친구가 나왔다.

90. 가위로 아영이가 억지로 문을 열고 집으로 몰래 들어왔다.

91. 따뜻한 커피를 좋아하는 여자가 매일 아침 커피를 마신다.

92. 운동을 좋아하는 남자가 친구들과 매일 오후 농구를 한다.

93. 일요일에 형이 집에 와서 엄마를 도왔다고 수호가 말했다.

94. 아침에 언니가 아빠와 차 안에서 통화했다고 미정이가 말했다.

95. 은희에게 아버지가 백화점에서 예쁜 옷과 인형을 사 주었다.

96. 지훈이에게 친구들이 편지를 써서 선물과 함께 전해 주었다.

97. 현우는 솔직히 오늘 산 옷이 마음에 들지 않았다.

98. 지은이는 열심히 공부해서 이번 시험에 좋은 성적으로 합격했다.

99. 칼로 엄마가 먹기 좋게 사과를 깎아서 아이들에게 주었다.

100. 연필로 수미가 친구의 얼굴을 그려서 생일 선물로 주었다.

101. 미래를 보는 아이가 가족에게 내일 일을 이야기해 주었다.

102. 물을 싫어하는 남자가 평생 한 번도 씻지 않았다.

103. 부자인 사장이 아무도 모르게 재산을 모두 학교에 기부했다.

104. 사랑에 빠진 그 여자가 1주일 만에 결혼을 결심했다.

105. 지훈이에게 선생님이 시험 문제의 답을 꿈속에서 알려 주었다.

106. 은주에게 지은이가 하늘을 나는 모습을 실제로 보여 주었다.

107. 혜진이는 어떤 책도 한 번 읽으면 다 외웠다.

108. 서현이는 놀랍게도 내일 일어날 일을 미리 알고 있었다.

109. 컴퓨터로 민지가 어제 보낸 서류가 갑자기 다 지워졌다.

110. 빨대로 민호는 뜨거운 커피도 한 번에 다 마셨다.

111. 남자가 싫어하는 그 여자를 친구들이 모르고 파티에 초대했다.

112. 개가 좋아하는 그 돼지를 주인이 어제 친구에게 팔았다.

113. 공원에서 할아버지가 할머니와 다정하게 손잡고 산책했다고 채린이가 말했다.

114. 방에서 아버지가 어머니와 웃으며 즐겁게 이야기했다고 영철이가 말했다.

115. 친구들에게 현우가 마침내 그 시험의 합격 소식을 알렸다.

116. 유진이에게 민지가 어제 본 드라마의 내용을 들려 주었다.

117. 동현이는 숙제를 먼저 끝내고 친구들과 농구를 하기로 했다.

118. 민아는 읽던 책을 끝까지 다 읽고 영화를 보았다.

119. 약으로 의사가 병을 고쳤고 아픈 환자들이 많이 줄어들었다.

120. 비누로 민호가 얼굴을 깨끗이 씻고 수건으로 얼른 닦았다.

Experiment 3

Target sentences

Condition a: New-given context & canonical word order

Condition b: New-given context & scrambled word order

Condition c: Given-new context & canonical word order

Condition d: Given-new context & scrambled word order

1a. Context	응급실에 환자가 한 명 남아 있었다.
1a. Target	어떤 의사가 그 환자를 정성껏 치료했다고 철수가 생각했다.
1b. Context	응급실에 의사가 한 명 남아 있었다.
1b. Target	어떤 환자를 그 의사가 정성껏 치료했다고 철수가 생각했다.
1c. Context	응급실에 의사가 한 명 남아 있었다.
1c. Target	그 의사가 어떤 환자를 정성껏 치료했다고 철수가 생각했다.
1d. Context	응급실에 환자가 한 명 남아 있었다.
1d. Target	그 환자를 어떤 의사가 정성껏 치료했다고 철수가 생각했다.
2a. Context	아시아 국가 모임에 대통령이 한 명 앉아 있었다.
2a. Target	어떤 화가가 그 대통령을 몰래 그렸다고 영희가 생각했다.
2b. Context	아시아 국가 모임에 화가가 한 명 앉아 있었다.
2b. Target	어떤 대통령을 그 화가가 몰래 그렸다고 영희가 생각했다.
2c. Context	아시아 국가 모임에 화가가 한 명 앉아 있었다.
2c. Target	그 화가가 어떤 대통령을 몰래 그렸다고 영희가 생각했다.
2d. Context	아시아 국가 모임에 대통령이 한 명 앉아 있었다.
2d. Target	그 대통령을 어떤 화가가 몰래 그렸다고 영희가 생각했다.
3a. Context	골목길에 도둑이 한 명 뛰어가고 있었다.

3a. Target	어떤 경찰이 그 도둑을 결국 잡았다고 정아가 생각했다.
3b. Context	골목길에 경찰이 한 명 뛰어가고 있었다.
3b. Target	어떤 도둑을 그 경찰이 결국 잡았다고 정아가 생각했다.
3c. Context	골목길에 경찰이 한 명 뛰어가고 있었다.
3c. Target	그 경찰이 어떤 도둑을 결국 잡았다고 정아가 생각했다.
3d. Context	골목길에 도둑이 한 명 뛰어가고 있었다.
3d. Target	그 도둑을 어떤 경찰이 결국 잡았다고 정아가 생각했다.
4a. Context	드라마 속 장면에서 직원이 한 명 일하고 있었다.
4a. Target	어떤 사장님이 그 직원을 냉정하게 해고했다고 민호가 생각했다.
4b. Context	드라마 속 장면에서 사장님이 한 명 일하고 있었다.
4b. Target	어떤 직원을 그 사장님이 냉정하게 해고했다고 민호가 생각했다.
4c. Context	드라마 속 장면에서 사장님이 한 명 일하고 있었다.
4c. Target	그 사장님이 어떤 직원을 냉정하게 해고했다고 민호가 생각했다.
4d. Context	드라마 속 장면에서 직원이 한 명 일하고 있었다.
4d. Target	그 직원을 어떤 사장님이 냉정하게 해고했다고 민호가 생각했다.
5a. Context	식당에 아이가 한 명 들어왔다.
5a. Target	어떤 아저씨가 그 아이를 심하게 혼냈다고 철수가 생각했다.
5b. Context	식당에 아저씨가 한 명 들어왔다.
5b. Target	어떤 아이를 그 아저씨가 심하게 혼냈다고 철수가 생각했다.
5c. Context	식당에 아저씨가 한 명 들어왔다.
5c. Target	그 아저씨가 어떤 아이를 심하게 혼냈다고 철수가 생각했다.
5d. Context	식당에 아이가 한 명 들어왔다.

5d. Target	그 아이를 어떤 아저씨가 심하게 혼냈다고 철수가 생각했다.
6a. Context	길가에서 손님이 한 명 기다리고 있었다.
6a. Target	어떤 운전사가 그 손님을 친절하게 태웠다고 철수가 생각했다.
6b. Context	길가에서 운전사가 한 명 기다리고 있었다.
6b. Target	어떤 손님을 그 운전사가 친절하게 태웠다고 철수가 생각했다.
6c. Context	길가에서 운전사가 한 명 기다리고 있었다.
6c. Target	그 운전사가 어떤 손님을 친절하게 태웠다고 철수가 생각했다.
6d. Context	길가에서 손님이 한 명 기다리고 있었다.
6d. Target	그 손님을 어떤 운전사가 친절하게 태웠다고 철수가 생각했다.
7a. Context	비행기에 손님이 한 명 서 있었다.
7a. Target	어떤 스튜어디스가 그 손님을 천천히 안내했다고 민호가 생각했다.
7b. Context	비행기에 스튜어디스가 한 명 서 있었다.
7b. Target	어떤 손님을 그 스튜어디스가 천천히 안내했다고 민호가 생각했다.
7c. Context	비행기에 스튜어디스가 한 명 서 있었다.
7c. Target	그 스튜어디스가 어떤 손님을 천천히 안내했다고 민호가 생각했다.
7d. Context	비행기에 손님이 한 명 서 있었다.
7d. Target	그 손님을 어떤 스튜어디스가 천천히 안내했다고 민호가 생각했다.
8a. Context	뉴스에 군인이 한 명 나왔다.
8a. Target	어떤 기자가 그 군인을 짧게 인터뷰했다고 영희가 생각했다.
8b. Context	뉴스에 기자가 한 명 나왔다.
8b. Target	어떤 군인을 그 기자가 짧게 인터뷰했다고 영희가 생각했다.
8c. Context	뉴스에 기자가 한 명 나왔다.

8c. Target	그 기자가 어떤 군인을 짧게 인터뷰했다고 영희가 생각했다.
8d. Context	뉴스에 군인이 한 명 나왔다.
8d. Target	그 군인을 어떤 기자가 짧게 인터뷰했다고 영희가 생각했다.
9a. Context	영화 속 장면에 직원이 한 명 나왔다.
9a. Target	어떤 회장님이 그 직원을 크게 혼냈다고 영희가 생각했다.
9b. Context	영화 속 장면에 회장님이 한 명 나왔다.
9b. Target	어떤 직원을 그 회장님이 크게 혼냈다고 영희가 생각했다.
9c. Context	영화 속 장면에 회장님이 한 명 나왔다.
9c. Target	그 회장님이 어떤 직원을 크게 혼냈다고 영희가 생각했다.
9d. Context	영화 속 장면에 직원이 한 명 나왔다.
9d. Target	그 직원을 어떤 회장님이 크게 혼냈다고 영희가 생각했다.
10a. Context	병실에 환자가 한 명 들어왔다.
10a. Target	어떤 간호사가 그 환자를 열심히 돌봤다고 정아가 생각했다.
10b. Context	병실에 간호사가 한 명 들어왔다.
10b. Target	어떤 환자를 그 간호사가 열심히 돌봤다고 정아가 생각했다.
10c. Context	병실에 간호사가 한 명 들어왔다.
10c. Target	그 간호사가 어떤 환자를 열심히 돌봤다고 정아가 생각했다.
10d. Context	병실에 환자가 한 명 들어왔다.
10d. Target	그 환자를 어떤 간호사가 열심히 돌봤다고 정아가 생각했다.
11a. Context	운동장에 선수가 한 명 남아 있었다.
11a. Target	어떤 코치가 그 선수를 직접 지도했다고 영희가 생각했다.
11b. Context	운동장에 코치가 한 명 남아 있었다.

11b. Target 어떤 선수를 그 코치가 직접 지도했다고 영희가 생각했다.

11c. Context 운동장에 코치가 한 명 남아 있었다.

11c. Target 그 코치가 어떤 선수를 직접 지도했다고 영희가 생각했다.

11d. Context 운동장에 선수가 한 명 남아 있었다.

11d. Target 그 선수를 어떤 코치가 직접 지도했다고 영희가 생각했다.

12a. Context 공원에 모델이 한 명 기다리고 있었다.

12a. Target 어떤 사진사가 그 모델을 예쁘게 찍었다고 정아가 생각했다.

12b. Context 공원에 사진사가 한 명 기다리고 있었다.

12b. Target 어떤 모델을 그 사진사가 예쁘게 찍었다고 정아가 생각했다.

12c. Context 공원에 사진사가 한 명 기다리고 있었다.

12c. Target 그 사진사가 어떤 모델을 예쁘게 찍었다고 정아가 생각했다.

12d. Context 공원에 모델이 한 명 기다리고 있었다.

12d. Target 그 모델을 어떤 사진사가 예쁘게 찍었다고 정아가 생각했다.

13a. Context 불 속에서 아이가 한 명 소리치고 있었다.

13a. Target 어떤 소방관이 그 아이를 안전하게 구했다고 민호가 생각했다.

13b. Context 불 속에서 소방관이 한 명 소리치고 있었다.

13b. Target 어떤 아이를 그 소방관이 안전하게 구했다고 민호가 생각했다.

13c. Context 불 속에서 소방관이 한 명 소리치고 있었다.

13c. Target 그 소방관이 어떤 아이를 안전하게 구했다고 민호가 생각했다.

13d. Context 불 속에서 아이가 한 명 소리치고 있었다.

13d. Target 그 아이를 어떤 소방관이 안전하게 구했다고 민호가 생각했다.

14a. Context 추운 밤에 노인이 한 명 걷고 있었다.

14a. Target 어떤 군인이 그 노인을 끝까지 보호했다고 철수가 생각했다.

14b. Context 추운 밤에 군인이 한 명 걷고 있었다.

14b. Target 어떤 노인을 그 군인이 끝까지 보호했다고 철수가 생각했다.

14c. Context 추운 밤에 군인이 한 명 걷고 있었다.

14c. Target 그 군인이 어떤 노인을 끝까지 보호했다고 철수가 생각했다.

14d. Context 추운 밤에 노인이 한 명 걷고 있었다.

14d. Target 그 노인을 어떤 군인이 끝까지 보호했다고 철수가 생각했다.

15a. Context 교실에 학생이 한 명 있었다.

15a. Target 어떤 선생님이 그 학생을 조용히 가르쳤다고 철수가 생각했다.

15b. Context 교실에 선생님이 한 명 있었다.

15b. Target 어떤 학생을 그 선생님이 조용히 가르쳤다고 철수가 생각했다.

15c. Context 교실에 선생님이 한 명 있었다.

15c. Target 그 선생님이 어떤 학생을 조용히 가르쳤다고 철수가 생각했다.

15d. Context 교실에 학생이 한 명 있었다.

15d. Target 그 학생을 어떤 선생님이 조용히 가르쳤다고 철수가 생각했다.

16a. Context 한밤중에 도둑이 한 명 밖으로 뛰어나왔다.

16a. Target 어떤 이웃이 그 도둑을 전화로 신고했다고 영희가 생각했다.

16b. Context 한밤중에 이웃이 한 명 밖으로 뛰어나왔다.

16b. Target 어떤 도둑을 그 이웃이 전화로 신고했다고 영희가 생각했다.

16c. Context 한밤중에 이웃이 한 명 밖으로 뛰어나왔다.

16c. Target 그 이웃이 어떤 도둑을 전화로 신고했다고 영희가 생각했다.

16d. Context 한밤중에 도둑이 한 명 밖으로 뛰어나왔다.

16d. Target 그 도둑을 어떤 이웃이 전화로 신고했다고 영희가 생각했다.

17a. Context 시골에 돼지가 한 마리 살았다.

17a. Target 어떤 남자가 그 돼지를 특별하게 키웠다고 민호가 생각했다.

17b. Context 시골에 남자가 한 명 살았다.

17b. Target 어떤 돼지를 그 남자가 특별하게 키웠다고 민호가 생각했다.

17c. Context 시골에 남자가 한 명 살았다.

17c. Target 그 남자가 어떤 돼지를 특별하게 키웠다고 민호가 생각했다.

17d. Context 시골에 돼지가 한 마리 살았다.

17d. Target 그 돼지를 어떤 남자가 특별하게 키웠다고 민호가 생각했다.

18a. Context 창가에 고양이가 한 마리 앉아 있었다.

18a. Target 어떤 여자가 그 고양이를 다정하게 쓰다듬었다고 정아가 생각했다.

18b. Context 창가에 여자가 한 명 앉아 있었다.

18b. Target 어떤 고양이를 그 여자가 다정하게 쓰다듬었다고 정아가 생각했다.

18c. Context 창가에 여자가 한 명 앉아 있었다.

18c. Target 그 여자가 어떤 고양이를 다정하게 쓰다듬었다고 정아가 생각했다.

18d. Context 창가에 고양이가 한 마리 앉아 있었다.

18d. Target 그 고양이를 어떤 여자가 다정하게 쓰다듬었다고 정아가 생각했다.

19a. Context 바닷가에 거북이가 한 마리 있었다.

19a. Target 어떤 아이가 그 거북이를 힘없이 놓아줬다고 정아가 생각했다.

19b. Context 바닷가에 아이가 한 명 있었다.

19b. Target 어떤 거북이를 그 아이가 힘없이 놓아줬다고 정아가 생각했다.

19c. Context 바닷가에 아이가 한 명 있었다.

19c. Target	그 아이가 어떤 거북이를 힘없이 놓아줬다고 정아가 생각했다.

19d. Context	바닷가에 거북이가 한 마리 있었다.

19d. Target	그 거북이를 어떤 아이가 힘없이 놓아줬다고 정아가 생각했다.

20a. Context	놀이터에 강아지가 한 마리 있었다.

20a. Target	어떤 남자아이가 그 강아지를 조심스럽게 안았다고 영희가 생각했다.

20b. Context	놀이터에 남자아이가 한 명 있었다.

20b. Target	어떤 강아지를 그 남자아이가 조심스럽게 안았다고 영희가 생각했다.

20c. Context	놀이터에 남자아이가 한 명 있었다.

20c. Target	그 남자아이가 어떤 강아지를 조심스럽게 안았다고 영희가 생각했다.

20d. Context	놀이터에 강아지가 한 마리 있었다.

20d. Target	그 강아지를 어떤 남자아이가 조심스럽게 안았다고 영희가 생각했다.

21a. Context	동물원에 코끼리가 한 마리 있었다.

21a. Target	어떤 할아버지가 그 코끼리를 제대로 길들였다고 철수가 생각했다.

21b. Context	동물원에 할아버지가 한 명 있었다.

21b. Target	어떤 코끼리를 그 할아버지가 제대로 길들였다고 철수가 생각했다.

21c. Context	동물원에 할아버지가 한 명 있었다.

21c. Target	그 할아버지가 어떤 코끼리를 제대로 길들였다고 철수가 생각했다.

21d. Context	동물원에 코끼리가 한 마리 있었다.

21d. Target	그 코끼리를 어떤 할아버지가 제대로 길들였다고 철수가 생각했다.

22a. Context	그늘 밑에 새가 한 마리 앉아 있었다.

22a. Target	어떤 아주머니가 그 새를 손으로 만졌다고 정아가 생각했다.

22b. Context	그늘 밑에 아주머니가 한 명 앉아 있었다.

22b. Target 어떤 새를 그 아주머니가 손으로 만졌다고 정아가 생각했다.

22c. Context 그늘 밑에 아주머니가 한 명 앉아 있었다.

22c. Target 그 아주머니가 어떤 새를 손으로 만졌다고 정아가 생각했다.

22d. Context 그늘 밑에 새가 한 마리 앉아 있었다.

22d. Target 그 새를 어떤 아주머니가 손으로 만졌다고 정아가 생각했다.

23a. Context 마을에 소가 한 마리 있었다.

23a. Target 어떤 할머니가 그 소를 혼자 돌봤다고 민호가 생각했다.

23b. Context 마을에 할머니가 한 명 있었다.

23b. Target 어떤 소를 그 할머니가 혼자 돌봤다고 민호가 생각했다.

23c. Context 마을에 할머니가 한 명 있었다.

23c. Target 그 할머니가 어떤 소를 혼자 돌봤다고 민호가 생각했다.

23d. Context 마을에 소가 한 마리 있었다.

23d. Target 그 소를 어떤 할머니가 혼자 돌봤다고 민호가 생각했다.

24a. Context 풀 속에 사자가 한 마리 숨어 있었다.

24a. Target 어떤 화가가 그 사자를 재빨리 그렸다고 정아가 생각했다.

24b. Context 풀 속에 화가가 한 명 숨어 있었다.

24b. Target 어떤 사자를 그 화가가 재빨리 그렸다고 정아가 생각했다.

24c. Context 풀 속에 화가가 한 명 숨어 있었다.

24c. Target 그 화가가 어떤 사자를 재빨리 그렸다고 정아가 생각했다.

24d. Context 풀 속에 사자가 한 마리 숨어 있었다.

24d. Target 그 사자를 어떤 화가가 재빨리 그렸다고 정아가 생각했다.

25a. Context 산에서 호랑이가 한 마리 내려오고 있었다.

25a. Target 어떤 여행자가 그 호랑이를 간신히 피했다고 영희가 생각했다.

25b. Context 산에서 여행자가 한 명 내려오고 있었다.

25b. Target 어떤 호랑이를 그 여행자가 간신히 피했다고 영희가 생각했다.

25c. Context 산에서 여행자가 한 명 내려오고 있었다.

25c. Target 그 여행자가 어떤 호랑이를 간신히 피했다고 영희가 생각했다.

25d. Context 산에서 호랑이가 한 마리 내려오고 있었다.

25d. Target 그 호랑이를 어떤 여행자가 간신히 피했다고 영희가 생각했다.

26a. Context 시장에 토끼가 한 마리 있었다.

26a. Target 어떤 고등학생이 그 토끼를 비싸게 샀다고 민호가 생각했다.

26b. Context 시장에 고등학생이 한 명 있었다.

26b. Target 어떤 토끼를 그 고등학생이 비싸게 샀다고 민호가 생각했다.

26c. Context 시장에 고등학생이 한 명 있었다.

26c. Target 그 고등학생이 어떤 토끼를 비싸게 샀다고 민호가 생각했다.

26d. Context 시장에 토끼가 한 마리 있었다.

26d. Target 그 토끼를 어떤 고등학생이 비싸게 샀다고 민호가 생각했다.

27a. Context 꽃밭에 나비가 한 마리 앉아 있었다.

27a. Target 어떤 여자아이가 그 나비를 어렵게 찾았다고 철수가 생각했다.

27b. Context 꽃밭에 여자아이가 한 명 앉아 있었다.

27b. Target 어떤 나비를 그 여자아이가 어렵게 찾았다고 철수가 생각했다.

27c. Context 꽃밭에 여자아이가 한 명 앉아 있었다.

27c. Target 그 여자아이가 어떤 나비를 어렵게 찾았다고 철수가 생각했다.

27d. Context 꽃밭에 나비가 한 마리 앉아 있었다.

27d. Target 그 나비를 어떤 여자아이가 어렵게 찾았다고 철수가 생각했다.

28a. Context 잔디밭에 비둘기가 한 마리 있었다.

28a. Target 어떤 대학생이 그 비둘기를 극적으로 구했다고 영희가 생각했다.

28b. Context 잔디밭에 대학생이 한 명 있었다.

28b. Target 어떤 비둘기를 그 대학생이 극적으로 구했다고 영희가 생각했다.

28c. Context 잔디밭에 대학생이 한 명 있었다.

28c. Target 그 대학생이 어떤 비둘기를 극적으로 구했다고 영희가 생각했다.

28d. Context 잔디밭에 비둘기가 한 마리 있었다.

28d. Target 그 비둘기를 어떤 대학생이 극적으로 구했다고 영희가 생각했다.

29a. Context 산속에 원숭이가 한 마리 살고 있었다.

29a. Target 어떤 노인이 그 원숭이를 잠시 길렀다고 영희가 생각했다.

29b. Context 산속에 노인이 한 명 살고 있었다.

29b. Target 어떤 원숭이를 그 노인이 잠시 길렀다고 영희가 생각했다.

29c. Context 산속에 노인이 한 명 살고 있었다.

29c. Target 그 노인이 어떤 원숭이를 잠시 길렀다고 영희가 생각했다.

29d. Context 산속에 원숭이가 한 마리 살고 있었다.

29d. Target 그 원숭이를 어떤 노인이 잠시 길렀다고 영희가 생각했다.

30a. Context 텔레비전에 물고기가 한 마리 나왔다.

30a. Target 어떤 초등학생이 그 물고기를 자세히 관찰했다고 민호가 생각했다.

30b. Context 텔레비전에 초등학생이 한 명 나왔다.

30b. Target 어떤 물고기를 그 초등학생이 자세히 관찰했다고 민호가 생각했다.

30c. Context 텔레비전에 초등학생이 한 명 나왔다.

30c. Target	그 초등학생이 어떤 물고기를 자세히 관찰했다고 민호가 생각했다.

30d. Context	텔레비전에 물고기가 한 마리 나왔다.

30d. Target	그 물고기를 어떤 초등학생이 자세히 관찰했다고 민호가 생각했다.

31a. Context	애완동물 가게에 병아리가 한 마리 있었다.

31a. Target	어떤 누나가 그 병아리를 사랑으로 보살폈다고 철수가 생각했다.

31b. Context	애완동물 가게에 누나가 한 명 있었다.

31b. Target	어떤 병아리를 그 누나가 사랑으로 보살폈다고 철수가 생각했다.

31c. Context	애완동물 가게에 누나가 한 명 있었다.

31c. Target	그 누나가 어떤 병아리를 사랑으로 보살폈다고 철수가 생각했다.

31d. Context	애완동물 가게에 병아리가 한 마리 있었다.

31d. Target	그 병아리를 어떤 누나가 사랑으로 보살폈다고 철수가 생각했다.

32a. Context	마당에서 말이 한 마리 기다리고 있었다.

32a. Target	어떤 소년이 그 말을 서투르게 탔다고 민호가 생각했다.

32b. Context	마당에서 소년이 한 명 기다리고 있었다.

32b. Target	어떤 말을 그 소년이 서투르게 탔다고 민호가 생각했다.

32c. Context	마당에서 소년이 한 명 기다리고 있었다.

32c. Target	그 소년이 어떤 말을 서투르게 탔다고 민호가 생각했다.

32d. Context	마당에서 말이 한 마리 기다리고 있었다.

32d. Target	그 말을 어떤 소년이 서투르게 탔다고 민호가 생각했다.

33a. Context	숲 속에 양이 한 마리 숨어 있었다.

33a. Target	어떤 늑대가 그 양을 계속 기다렸다고 영희가 생각했다.

33b. Context	숲 속에 늑대가 한 마리 숨어 있었다.

33b. Target	어떤 양을 그 늑대가 계속 기다렸다고 영희가 생각했다.
33c. Context	숲 속에 늑대가 한 마리 숨어 있었다.
33c. Target	그 늑대가 어떤 양을 계속 기다렸다고 영희가 생각했다.
33d. Context	숲 속에 양이 한 마리 숨어 있었다.
33d. Target	그 양을 어떤 늑대가 계속 기다렸다고 영희가 생각했다.
34a. Context	어둠 속에 토끼가 한 마리 있었다.
34a. Target	어떤 사자가 그 토끼를 거칠게 잡았다고 정아가 생각했다.
34b. Context	어둠 속에 사자가 한 마리 있었다.
34b. Target	어떤 토끼를 그 사자가 거칠게 잡았다고 정아가 생각했다.
34c. Context	어둠 속에 사자가 한 마리 있었다.
34c. Target	그 사자가 어떤 토끼를 거칠게 잡았다고 정아가 생각했다.
34d. Context	어둠 속에 토끼가 한 마리 있었다.
34d. Target	그 토끼를 어떤 사자가 거칠게 잡았다고 정아가 생각했다.
35a. Context	바위 밑에 쥐가 한 마리 있었다.
35a. Target	어떤 뱀이 그 쥐를 한입에 먹었다고 철수가 생각했다.
35b. Context	바위 밑에 뱀이 한 마리 있었다.
35b. Target	어떤 쥐를 그 뱀이 한입에 먹었다고 철수가 생각했다.
35c. Context	바위 밑에 뱀이 한 마리 있었다.
35c. Target	그 뱀이 어떤 쥐를 한입에 먹었다고 철수가 생각했다.
35d. Context	바위 밑에 쥐가 한 마리 있었다.
35d. Target	그 쥐를 어떤 뱀이 한입에 먹었다고 철수가 생각했다.
36a. Context	집 앞에 새가 한 마리 앉아 있었다.

36a. Target　어떤 고양이가 그 새를 순식간에 붙잡았다고 민호가 생각했다.

36b. Context　집 앞에 고양이가 한 마리 앉아 있었다.

36b. Target　어떤 새를 그 고양이가 순식간에 붙잡았다고 민호가 생각했다.

36c. Context　집 앞에 고양이가 한 마리 앉아 있었다.

36c. Target　그 고양이가 어떤 새를 순식간에 붙잡았다고 민호가 생각했다.

36d. Context　집 앞에 새가 한 마리 앉아 있었다.

36d. Target　그 새를 어떤 고양이가 순식간에 붙잡았다고 민호가 생각했다.

37a. Context　물가에 말이 한 마리 있었다.

37a. Target　어떤 악어가 그 말을 세게 물었다고 영희가 생각했다.

37b. Context　물가에 악어가 한 마리 있었다.

37b. Target　어떤 말을 그 악어가 세게 물었다고 영희가 생각했다.

37c. Context　물가에 악어가 한 마리 있었다.

37c. Target　그 악어가 어떤 말을 세게 물었다고 영희가 생각했다.

37d. Context　물가에 말이 한 마리 있었다.

37d. Target　그 말을 어떤 악어가 세게 물었다고 영희가 생각했다.

38a. Context　이웃집에 오리가 한 마리 있었다.

38a. Target　어떤 개가 그 오리를 갑자기 공격했다고 정아가 생각했다.

38b. Context　이웃집에 개가 한 마리 있었다.

38b. Target　어떤 오리를 그 개가 갑자기 공격했다고 정아가 생각했다.

38c. Context　이웃집에 개가 한 마리 있었다.

38c. Target　그 개가 어떤 오리를 갑자기 공격했다고 정아가 생각했다.

38d. Context　이웃집에 오리가 한 마리 있었다.

38d. Target	그 오리를 어떤 개가 갑자기 공격했다고 정아가 생각했다.
39a. Context	수족관에서 물고기가 한 마리 헤엄치고 있었다.
39a. Target	어떤 펭귄이 그 물고기를 그대로 삼켰다고 철수가 생각했다.
39b. Context	수족관에서 펭귄이 한 마리 헤엄치고 있었다.
39b. Target	어떤 물고기를 그 펭귄이 그대로 삼켰다고 철수가 생각했다.
39c. Context	수족관에서 펭귄이 한 마리 헤엄치고 있었다.
39c. Target	그 펭귄이 어떤 물고기를 그대로 삼켰다고 철수가 생각했다.
39d. Context	수족관에서 물고기가 한 마리 헤엄치고 있었다.
39d. Target	그 물고기를 어떤 펭귄이 그대로 삼켰다고 철수가 생각했다.
40a. Context	나무 밑에 원숭이가 한 마리 있었다.
40a. Target	어떤 호랑이가 그 원숭이를 쉽게 사냥했다고 민호가 생각했다.
40b. Context	나무 밑에 호랑이가 한 마리 있었다.
40b. Target	어떤 원숭이를 그 호랑이가 쉽게 사냥했다고 민호가 생각했다.
40c. Context	나무 밑에 호랑이가 한 마리 있었다.
40c. Target	그 호랑이가 어떤 원숭이를 쉽게 사냥했다고 민호가 생각했다.
40d. Context	나무 밑에 원숭이가 한 마리 있었다.
40d. Target	그 원숭이를 어떤 호랑이가 쉽게 사냥했다고 민호가 생각했다.
41a. Context	빗속에서 병아리가 한 마리 걷고 있었다.
41a. Target	어떤 닭이 그 병아리를 철저히 보호했다고 영희가 생각했다.
41b. Context	빗속에서 닭이 한 마리 걷고 있었다.
41b. Target	어떤 병아리를 그 닭이 철저히 보호했다고 영희가 생각했다.
41c. Context	빗속에서 닭이 한 마리 걷고 있었다.

41c. Target	그 닭이 어떤 병아리를 철저히 보호했다고 영희가 생각했다.
41d. Context	빗속에서 병아리가 한 마리 걷고 있었다.
41d. Target	그 병아리를 어떤 닭이 철저히 보호했다고 영희가 생각했다.
42a. Context	풀밭에 소가 한 마리 나타났다.
42a. Target	어떤 사자가 그 소를 곧바로 따라갔다고 정아가 생각했다.
42b. Context	풀밭에 사자가 한 마리 나타났다.
42b. Target	어떤 소를 그 사자가 곧바로 따라갔다고 정아가 생각했다.
42c. Context	풀밭에 사자가 한 마리 나타났다.
42c. Target	그 사자가 어떤 소를 곧바로 따라갔다고 정아가 생각했다.
42d. Context	풀밭에 소가 한 마리 나타났다.
42d. Target	그 소를 어떤 사자가 곧바로 따라갔다고 정아가 생각했다.
43a. Context	바닷속에 물고기가 한 마리 있었다.
43a. Target	어떤 고래가 그 물고기를 아깝게 놓쳤다고 철수가 생각했다.
43b. Context	바닷속에 고래가 한 마리 있었다.
43b. Target	어떤 물고기를 그 고래가 아깝게 놓쳤다고 철수가 생각했다.
43c. Context	바닷속에 고래가 한 마리 있었다.
43c. Target	그 고래가 어떤 물고기를 아깝게 놓쳤다고 철수가 생각했다.
43d. Context	바닷속에 물고기가 한 마리 있었다.
43d. Target	그 물고기를 어떤 고래가 아깝게 놓쳤다고 철수가 생각했다.
44a. Context	나뭇가지 위에 벌레가 한 마리 있었다.
44a. Target	어떤 원숭이가 그 벌레를 처음으로 찾았다고 민호가 생각했다.
44b. Context	나뭇가지 위에 원숭이가 한 마리 있었다.

44b. Target 어떤 벌레를 그 원숭이가 처음으로 찾았다고 민호가 생각했다.

44c. Context 나뭇가지 위에 원숭이가 한 마리 있었다.

44c. Target 그 원숭이가 어떤 벌레를 처음으로 찾았다고 민호가 생각했다.

44d. Context 나뭇가지 위에 벌레가 한 마리 있었다.

44d. Target 그 벌레를 어떤 원숭이가 처음으로 찾았다고 민호가 생각했다.

45a. Context 농장에 닭이 한 마리 살고 있었다.

45a. Target 어떤 돼지가 그 닭을 억지로 밀어냈다고 정아가 생각했다.

45b. Context 농장에 돼지가 한 마리 살고 있었다.

45b. Target 어떤 닭을 그 돼지가 억지로 밀어냈다고 정아가 생각했다.

45c. Context 농장에 돼지가 한 마리 살고 있었다.

45c. Target 그 돼지가 어떤 닭을 억지로 밀어냈다고 정아가 생각했다.

45d. Context 농장에 닭이 한 마리 살고 있었다.

45d. Target 그 닭을 어떤 돼지가 억지로 밀어냈다고 정아가 생각했다.

46a. Context 논에 오리가 한 마리 서 있었다.

46a. Target 어떤 소가 그 오리를 실수로 찼다고 정아가 생각했다.

46b. Context 논에 소가 한 마리 서 있었다.

46b. Target 어떤 오리를 그 소가 실수로 찼다고 정아가 생각했다.

46c. Context 논에 소가 한 마리 서 있었다.

46c. Target 그 소가 어떤 오리를 실수로 찼다고 정아가 생각했다.

46d. Context 논에 오리가 한 마리 서 있었다.

46d. Target 그 오리를 어떤 소가 실수로 찼다고 정아가 생각했다.

47a. Context 집에 양이 한 마리 있었다.

47a. Target 어떤 개가 그 양을 밤새도록 지켰다고 철수가 생각했다.

47b. Context 집에 개가 한 마리 있었다.

47b. Target 어떤 양을 그 개가 밤새도록 지켰다고 철수가 생각했다.

47c. Context 집에 개가 한 마리 있었다.

47c. Target 그 개가 어떤 양을 밤새도록 지켰다고 철수가 생각했다.

47d. Context 집에 양이 한 마리 있었다.

47d. Target 그 양을 어떤 개가 밤새도록 지켰다고 철수가 생각했다.

48a. Context 지붕 위에 쥐가 한 마리 있었다.

48a. Target 어떤 고양이가 그 쥐를 재미로 쫓았다고 민호가 생각했다.

48b. Context 지붕 위에 고양이가 한 마리 있었다.

48b. Target 어떤 쥐를 그 고양이가 재미로 쫓았다고 민호가 생각했다.

48c. Context 지붕 위에 고양이가 한 마리 있었다.

48c. Target 그 고양이가 어떤 쥐를 재미로 쫓았다고 민호가 생각했다.

48d. Context 지붕 위에 쥐가 한 마리 있었다.

48d. Target 그 쥐를 어떤 고양이가 재미로 쫓았다고 민호가 생각했다.

Fillers

1. Context	노란 옷을 입은 여자가 가게로 들어왔다.
1. Target	이 여자를 본 남자가 손을 흔들었다.
2. Context	모자를 쓴 여자가 웃으며 뛰어왔다.
2. Target	이 여자를 좋아하는 남자가 꽃을 선물했다.
3. Context	남자가 약속 시간에 조금 늦었다.
3. Target	이 남자를 싫어하는 여자가 화를 냈다.
4. Context	돼지가 팔리지 않아 집으로 돌아왔다.
4. Target	이 돼지를 본 개가 꼬리를 흔들었다.
5. Context	어떤 남자가 도둑을 잡았다.
5. Target	이 남자가 잡은 도둑을 경찰이 체포했다.
6. Context	유명한 가수가 콘서트를 열었다.
6. Target	이 가수가 부른 노래를 팬들이 좋아했다.
7. Context	철수와 친구가 점심을 먹으러 나왔다.
7. Target	저 식당의 음식이 맛있다고 철수가 말했다.
8. Context	경찰의 질문에 민호가 대답하고 있었다.
8. Target	저 사람들이 서로 때렸다고 민호가 말했다.
9. Context	영희가 친구와 영화를 보러 갔다.
9. Target	저 극장에서 보자고 친구가 영희에게 말했다.
10. Context	어떤 고등학생이 열심히 공부했다.

10. Target	이 고등학생이 결국 좋은 대학교에 입학했다.
11. Context	어떤 남자가 회사에서 열심히 일했다.
11. Target	이 남자는 회사에 매일 일찍 출근했다.
12. Context	성실한 학생이 있었다.
12. Target	이 학생은 한 번도 지각하지 않았다.
13. Context	청소를 좋아하는 주부가 있었다.
13. Target	이 주부가 매일 집안을 깨끗하게 청소했다.
14. Context	대학생이 발표 준비를 하고 있었다.
14. Target	이 대학생이 컴퓨터가 갑자기 꺼져서 당황했다.
15. Context	영희와 정아가 텔레비전을 보고 있었다.
15. Target	이 둘이 드라마를 보다가 슬퍼서 울었다.
16. Context	민호는 서점에서 일했다.
16. Target	이 서점에서 책과 잡지를 사람들에게 팔았다.
17. Context	철수가 혼자 여행을 떠났다.
17. Target	이 여행에서 철수는 좋은 사람들을 만났다.
18. Context	정아의 아버지가 새로 핸드폰을 샀다.
18. Target	이 핸드폰으로 아버지가 정아와 길게 통화했다.
19. Context	엄마가 철수에게 오늘 계획을 물었다.
19. Target	저 방을 깨끗하게 청소하겠다고 철수가 말했다.
20. Context	지난 일요일에는 누나가 엄마를 도왔다.

20. Target	이번 일요일에는 형이 바쁜 엄마를 도와주었다.
21. Context	영희가 언니와 통화를 하고 있었다.
21. Target	언니 선물로 가방을 샀다고 영희가 말했다.
22. Context	친구가 철수에게 주말에 한 일을 물었다.
22. Target	집에서 계속 텔레비전을 봤다고 철수가 말했다.
23. Context	토요일은 집안일을 하기로 약속했다.
23. Target	철수가 청소를 하고 영희가 빨래를 했다.
24. Context	영희가 어제 늦게까지 공부를 했다.
24. Target	늦잠을 잤지만 영희는 학교에 늦지 않았다.
25. Context	이번 주부터 중간고사 기간이다.
25. Target	정아는 도서관에 갔지만 철수는 집에서 놀았다.
26. Context	영희가 점심을 먹으러 식당에 갔다.
26. Target	먹고 싶은 음식이 없어서 한참 고민했다.
27. Context	정아는 아침에 일찍 일어난다.
27. Target	일어나서 아침 운동을 하고 학교에 간다.
28. Context	아기가 아파서 울었다.
28. Target	아기의 엄마가 걱정이 되서 병원에 데려갔다.
29. Context	민호와 정아가 백화점에 갔다.
29. Target	친구의 생일 선물을 둘이 같이 샀다.
30. Context	민호가 친구와 이야기를 하고 있었다.

30. Target	저기 있는 남자를 안다고 민호가 말했다.
31. Context	정아가 오빠에게 전화를 걸었다.
31. Target	누군가 오빠에게 편지를 보냈다고 정아가 말했다.
32. Context	어린아이가 자동차 안에 혼자 있었다.
32. Target	저기 어린아이가 혼자 있다고 민호가 말했다.
33. Context	정아가 새로 생긴 가게에 갔다.
33. Target	거기서 제일 비싼 음식을 시켜서 먹었다.
34. Context	정아가 동료와 저녁을 먹기로 했다.
34. Target	저쪽에 좋은 식당이 많다고 정아가 말했다.
35. Context	민호가 친구들에게 화가 났다.
35. Target	민호가 싫어하는 여자를 친구들이 초대했기 때문이다.
36. Context	강아지가 하루 종일 밥을 먹지 않았다.
36. Target	강아지가 좋아하는 돼지를 주인이 팔았기 때문이다.
37. Context	어떤 여자가 사무실로 들어왔다.
37. Target	이 여자가 준비한 서류를 사람들이 칭찬했다.
38. Context	정아가 남자 친구가 생겼다.
38. Target	정아가 만나는 이 남자를 친구들이 싫어했다.
39. Context	돼지 한 마리와 개 한 마리가 있었다.
39. Target	돼지가 따라다니는 이 개를 어린아이가 안았다.
40. Context	아이가 엄마한테 혼이 났다.

40. Target	엄마가 혼낸 이 아이를 아빠가 위로했다.
41. Context	선생님이 영희를 칭찬했다.
41. Target	선생님이 칭찬한 영희를 친구들도 역시 좋아했다.
42. Context	주인이 강아지를 한 마리 안고 있었다.
42. Target	강아지를 좋아하는 돼지가 계속 그쪽만 쳐다봤다.
43. Context	엄마와 아이가 집에 있었다.
43. Target	숙제를 하던 아이가 엄마에게 질문을 했다.
44. Context	철수가 수영장에 갔다.
44. Target	물을 싫어하는 철수가 처음으로 수영을 배웠다.
45. Context	영희가 옷 가게에 갔다.
45. Target	빨간색을 좋아하는 영희가 빨간색 옷을 샀다.
46. Context	민호가 집에 혼자 있어서 심심했다.
46. Target	게임을 좋아하는 민호가 컴퓨터로 게임을 했다.
47. Context	신문에 어떤 도둑 이야기가 나왔다.
47. Target	물건을 훔친 도둑이 주인에게 잘못을 빌었다.
48. Context	오늘은 날씨가 아주 좋았다.
48. Target	오후에 민호와 철수가 만나서 운동을 했다.
49. Context	라디오에서 조용한 음악이 흘러나왔다.
49. Target	이 음악을 좋아하는 여자가 가만히 듣고 있었다.
50. Context	유명한 대학교에서 신입생을 모집했다.

50. Target	이 대학교에서 공부하고 싶었던 철수가 지원을 했다.
51. Context	친구가 야채로 간단히 요리를 했다.
51. Target	이런 요리를 좋아하는 영희가 맛있게 다 먹었다.
52. Context	어떤 남자가 길에서 돈을 주웠다.
52. Target	이 남자가 주운 돈을 아내가 은행에 저금했다.
53. Context	여자가 핸드폰으로 게임을 하고 있었다.
53. Target	이 게임은 어렵지만 재밌어서 요즘 인기가 많다.
54. Context	영희가 친구들과 노래를 불렀다.
54. Target	이 노래는 친구들이 요즘 좋아하는 가수의 노래이다.
55. Context	언니가 새를 한 마리 키웠다.
55. Target	이 새는 밤이 되면 혼자서 노래를 했다.
56. Context	영희가 실수로 거울을 깼다.
56. Target	이 깨진 거울은 친구가 선물한 소중한 물건이었다.
57. Context	어릴 때부터 여자가 자주 아팠다.
57. Target	이 여자가 매일매일 등산을 해서 지금은 건강해졌다.
58. Context	정아와 민호가 높은 산에 올라갔다.
58. Target	이 산에 올라가서 정아와 민호가 사진을 찍었다.
59. Context	철수의 누나가 빵을 사왔다.
59. Target	이 빵을 우유와 함께 철수가 맛있게 먹었다.
60. Context	영희가 아파서 병원에 갔다.

60. Target	이 병원에서 영희는 여러 가지 검사를 받았다.
61. Context	오늘은 중요한 시험이 있는 날이다.
61. Target	이 시험을 보러 영희도 오늘 학교에 갔다.
62. Context	민호의 핸드폰이 고장이 났다.
62. Target	이 핸드폰 대신 민호가 핸드폰을 새로 샀다.
63. Context	친구들이 같이 사진을 찍었다.
63. Target	이 사진을 친구들 모두 마음에 들어 했다.
64. Context	돈을 많이 번 사장님이 있었다.
64. Target	이 사장님이 자기 재산을 모두 학교에 기부했다.
65. Context	따뜻한 커피 두 잔이 있었다.
65. Target	이 커피 두 잔을 영희와 철수가 마셨다.
66. Context	민호가 강아지를 키웠다.
66. Target	이 강아지와 매일 아침 민호가 함께 걸었다.
67. Context	운동장에 친구들이 모여 있었다.
67. Target	이 친구들과 운동을 좋아하는 남자가 농구를 했다.
68. Context	철수가 식당에서 주문을 했다.
68. Target	점심으로 비빔밥과 만두를 시켜서 혼자 맛있게 먹었다.
69. Context	영희는 꽃과 나무를 좋아한다.
69. Target	취미로 큰 정원에 나무와 꽃을 많이 심었다.
70. Context	정아는 아침 식사를 꼭 한다.

70. Target	매일 정아가 커피와 빵을 아침으로 함께 먹었다.
71. Context	철수가 책 두 권을 선물로 받았다.
71. Target	집에서 철수가 그중에 한 권을 저녁에 읽었다.
72. Context	아침에 민호가 학교에 걸어갔다.
72. Target	가는 길에 우연히 친구를 만나서 같이 걸어갔다.
73. Context	비 오는 날 영희가 우산이 없었다.
73. Target	교실에서 걱정하는 영희에게 친구가 자기 우산을 주었다.
74. Context	정아가 배가 몹시 고팠다.
74. Target	냉장고에서 음식을 꺼내서 정아가 맛있게 모두 먹었다.
75. Context	철수의 동생이 엄마에게 혼나고 있었다.
75. Target	도서관에서 동생이 보던 새 책을 찢었기 때문이었다.
76. Context	정아가 감기에 걸렸다.
76. Target	감기에 걸려서 아팠지만 정아는 열심히 시험공부를 했다.
77. Context	어젯밤에 눈이 많이 왔다.
77. Target	눈이 와서 미끄러운 길을 철수가 천천히 운전했다.
78. Context	갑자기 날씨가 추워졌다.
78. Target	날씨가 너무 추워서 영희는 두꺼운 코트를 입었다.
79. Context	월급을 받지 못해서 민호가 돈이 없었다.
79. Target	돈이 없었지만 민호는 친구의 생일 선물을 샀다.
80. Context	영희가 어제 이사를 했다.

80. Target	영희는 너무 피곤해서 일에 집중할 수 없었다.
81. Context	민호와 정아가 만났다.
81. Target	민호가 지갑을 잃어버려서 정아에게 돈을 조금 빌렸다.
82. Context	여자가 길에서 첫사랑을 우연히 만났다.
82. Target	사랑에 빠진 여자가 그 사람과 결혼을 결심했다.
83. Context	정아의 아빠가 정아에게 선물을 주었다.
83. Target	아빠가 선물로 준 컴퓨터를 정아가 친구에게 팔았다.
84. Context	어떤 아이가 아버지를 돕고 있었다.
84. Target	아버지가 자른 나무를 이 아이가 모아서 옮겼다.
85. Context	영희의 엄마가 케이크를 만들었다.
85. Target	엄마가 만들어 준 케이크를 영희가 모두 먹었다.
86. Context	민호의 동생이 그림을 그렸다.
86. Target	동생이 그려 준 그림을 민호가 벽에 걸었다.
87. Context	철수의 친구가 이메일을 보냈다.
87. Target	친구가 보낸 이메일을 철수가 실수로 모두 지웠다.
88. Context	정아가 라디오를 듣고 있었다.
88. Target	좋아하는 가수가 부른 노래를 정아가 따라 불렀다.
89. Context	할머니가 목도리를 만들었다.
89. Target	할머니가 만들어 주신 목도리를 영희가 겨울에 사용했다.
90. Context	학교 앞에 새로운 빵집이 생겼다.

90. Target	맛있는 빵을 좋아하는 남자가 거기서 빵을 샀다.
91. Context	아침에 눈이 많이 쌓였다.
91. Target	눈을 좋아하는 민호가 누나와 눈사람을 크게 만들었다.
92. Context	지난달에 예쁜 꽃집이 문을 열었다.
92. Target	꽃을 좋아하는 여자가 매일 이 꽃집에 간다.
93. Context	집 근처에 큰 도서관이 생겼다.
93. Target	공부를 열심히 하는 학생이 매일 여기에 간다.
94. Context	테이블 위에 커피가 한 잔 놓여 있었다.
94. Target	커피를 매우 좋아하는 여자가 혼자 다 마셨다.
95. Context	정아가 친구를 집으로 초대했다.
95. Target	고기를 좋아하는 정아가 소고기로 요리를 맛있게 했다.
96. Context	영희와 철수가 어제 집에서 놀았다.
96. Target	영희와 철수는 거실에서 영화를 보면서 콜라를 마셨다.

Appendix B

Results of Experiments 1, 2, and 3

Experiment 1

Table A.1. Generalized linear mixed model results, agent identification task

	Offline Accuracy			
	Estimate	SE	*z*	*p*
NS				
(Intercept)	10.360	31.490	0.329	0.742
Plausibility	-9.359	62.922	-0.149	0.882
Word Order	-11.542	62.961	-0.183	0.855
Plausibility × Word Order	17.496	125.814	0.139	0.889
Random effects Structure	(1+Word Order||Subject) + (1+Plausibility||Item)			
L2				
(Intercept)	2.611246	0.250616	10.419	< 2e-16
Plausibility	0.706924	0.247825	2.853	0.00434
Word Order	0.262159	0.158206	1.657	0.09751
Cloze scores	0.115488	0.045873	2.518	0.01182
Plausibility × Word Order	0.673600	0.316097	2.131	0.03309
Plausibility × Cloze scores	-0.005137	0.039588	-0.130	0.89675
Word Order × Cloze scores	0.022361	0.032373	0.691	0.48974
Plausibility × Word Order × Cloze scores	0.002506	0.064564	0.039	0.96903
Random effects structure	(1+Plausibility|Subject) + (1|Item)			

Table A.2. Generalized linear mixed model results, probe recognition task

	Online Accuracy			
	Estimate	SE	*z*	*p*
NS				
(Intercept)	5.3198	0.7329	7.259	3.91e-13
Plausibility	0.2177	0.5911	0.368	0.713
Word Order	-0.1994	1.7428	-0.114	0.909
Plausibility × Word Order	-0.9874	1.1835	-0.834	0.404
Random effects structure	(1+Word Order||Subject) + (1 + Word Order||Item)			
L2				
(Intercept)	3.18276	0.25336	12.562	<2e-16
Plausibility	0.62202	0.32198	1.932	0.0534
Word Order	-0.20566	0.29799	-0.690	0.4901
Cloze scores	0.04308	0.03879	1.110	0.2668
Plausibility × Word Order	-0.19221	0.38903	-0.494	0.6212
Plausibility × Cloze scores	0.07329	0.04158	1.763	0.0780
Word Order × Cloze scores	0.07336	0.04098	1.790	0.0734
Plausibility × Word Order × Cloze scores	0.17655	0.08058	2.191	0.0284
Random effects structure	(1+Plausibility+Word Order||Subject) + (1+Plausibility+Word Order+Cloze scores||Item)			

Table A.3. Linear mixed effects model results for region 2

	Reading time			
	Estimate	SE	*t*	*p*
NS All trials				
(Intercept)	19.817	17.438	1.136	0.2675
Plausibility	-26.535	14.685	-1.807	0.0732
Word Order	8.363	15.556	0.538	0.5932
Plausibility × Word Order	-8.339	35.032	-0.238	0.8136
Random effects structure	(1+Plausibility*Word Order|Subject) +			
	(1+Plausibility+Word Order|Item)			
NS Correct trials				
(Intercept)	21.000	18.201	1.154	0.260
Plausibility	-21.249	15.309	-1.388	0.169
Word Order	5.537	16.679	0.332	0.742
Plausibility × Word Order	-10.687	33.894	-0.315	0.755
Random effects structure	(1+Plausibility * Word Order|Subject) +			
	(1+Plausibility+Word Order|Item)			
L2 All trials				
(Intercept)	407.2701	46.9333	8.678	5.01e-12
Plausibility	-28.9873	38.6947	-0.749	0.458750
Word Order	142.0102	38.8811	3.652	0.000819
Cloze scores	-11.2471	7.9313	-1.418	0.163583
WMC	-6.1372	5.4064	-1.135	0.263811

Table A.3 (continued). Linear mixed effects model results for region 2

	Reading time			
	Estimate	SE	*t*	*p*
Plausibility × Word Order	207.9512	66.5808	3.123	0.001818
Plausibility × Cloze scores	-5.1257	7.3119	-0.701	0.487947
Word Order × Cloze scores	6.7643	7.4340	0.910	0.368843
Plausibility × WMC	-1.7746	5.4175	-0.328	0.745089
Word Order × WMC	-12.3239	5.2823	-2.333	0.025383
Cloze scores × WMC	1.8946	1.0234	1.851	0.072354
Plausibility × Word Order × Cloze scores	-2.9032	13.0068	-0.223	0.823401
Plausibility × Word Order × WMC	-2.5724	9.2523	-0.278	0.781021
Plausibility × Cloze scores × WMC	0.9969	1.0218	0.976	0.335666
Word Order × Cloze scores × WMC	2.3722	1.0047	2.361	0.023695
Plausibility × Word Order × Cloze scores × WMC	1.9170	1.7576	1.091	0.275559
Random effects structure	(1+Plausibility+Word Order|Subject) + (1+Plausibility+Word Order+Cloze scores|Item)			
L2 Correct trials				
(Intercept)	387.2281	45.5725	8.497	1.34e-11

Table A.3 (continued). Linear mixed effects model results for region 2

	Reading time			
	Estimate	SE	*t*	*p*
Plausibility	-13.4315	39.4549	-0.340	0.73423
Word Order	102.4648	44.1183	2.322	0.02563
Cloze scores	-12.2066	7.4692	-1.634	0.11080
WMC	-3.3999	5.3032	-0.641	0.52538
Plausibility × Word Order	217.0229	75.1249	2.889	0.00393
Plausibility × Cloze scores	-6.0471	7.3552	-0.822	0.41250
Word Order × Cloze scores	6.5563	8.1711	0.802	0.42770
Plausibility × WMC	-4.4466	5.7454	-0.774	0.44001
Word Order × WMC	-11.9505	6.1956	-1.929	0.06025
Cloze scores × WMC	2.0782	0.9946	2.089	0.04379
Plausibility × Word Order × Cloze scores	-12.6798	14.2626	-0.889	0.37415
Plausibility × Word Order × WMC	-3.0816	10.8531	-0.284	0.77650
Plausibility × Cloze scores × WMC	1.2507	1.0535	1.187	0.23714
Word Order × Cloze scores × WMC	2.9548	1.1397	2.593	0.01322
Plausibility × Word Order × Cloze scores × WMC	2.9299	1.9977	1.467	0.14272

Table A.3 (continued). Linear mixed effects model results for region 2

	Reading time			
	Estimate	SE	*t*	*p*
Random effects structure	(1+Plausibility+Word Order|Subject) +			
	(1+Plausibility+Word Order+Cloze scores+WMC||Item)			

Table A.4. Linear mixed effects model results for region 3

	Reading time			
	Estimate	SE	*t*	*p*
NS All trials				
(Intercept)	109.31	34.38	3.179	0.00464
Plausibility	-69.82	34.08	-2.049	0.04838
Word Order	55.20	27.16	2.032	0.04992
Plausibility × Word Order	1.94	46.05	0.042	0.96641
Random effects structure	(1+Plausibility+Word Order|Subject) +			
	(1+Plausibility+Word Order|Item)			
NS Correct trials				
(Intercept)	113.17	35.12	3.222	0.00413
Plausibility	-70.13	33.77	-2.077	0.04529
Word Order	61.03	26.87	2.272	0.02953
Plausibility × Word Order	-13.00	47.08	-0.276	0.78258
Random effects structure	(1+Plausibility+Word Order|Subject) +			
	(1+Plausibility+Word Order||Item)			
L2 All trials				
(Intercept)	575.6325	65.6842	8.764	1.54e-12
Plausibility	-26.2651	42.0487	-0.625	0.535996
Word Order	-22.9558	42.6980	-0.538	0.594179
Cloze scores	3.7009	10.1877	0.363	0.718524
WMC	-1.9690	7.2872	-0.270	0.788547

Table A.4 (continued). Linear mixed effects model results for region 3

	Reading time			
	Estimate	SE	*t*	*p*
Plausibility × Word Order	180.2660	76.1379	2.368	0.018020
Plausibility × Cloze scores	10.9848	7.7696	1.414	0.161760
Word Order × Cloze scores	4.9076	8.2972	0.591	0.557784
Plausibility × WMC	-1.5711	5.7329	-0.274	0.784803
Word Order × WMC	8.5288	5.8810	1.450	0.155608
Cloze scores × WMC	0.2165	1.3788	0.157	0.876117
Plausibility × Word Order × Cloze scores	49.8441	14.8814	3.349	0.000828
Plausibility × Word Order × WMC	-16.9388	10.5484	-1.606	0.108504
Plausibility × Cloze scores × WMC	-0.7185	1.0581	-0.679	0.499259
Word Order × Cloze scores × WMC	-0.9237	1.1150	-0.828	0.412849
Plausibility × Word Order × Cloze scores × WMC	1.3858	1.9991	0.693	0.488291
Random effects structure	(1+Plausibility+Word Order|Subject) + (1+Plausibility+Word Order|Item)			
L2 Correct trials				
(Intercept)	584.7535	67.6043	8.650	3.07e-12

Table A.4 (continued). Linear mixed effects model results for region 3

	Reading time			
	Estimate	SE	*t*	*p*
Plausibility	-42.7308	50.4999	-0.846	0.4022
Word Order	-33.3778	52.5228	-0.635	0.5289
Cloze scores	2.8330	10.4296	0.272	0.7875
WMC	-1.2431	7.5483	-0.165	0.8701
Plausibility × Word Order	189.0264	87.4995	2.160	0.0309
Plausibility × Cloze scores	8.8922	9.1887	0.968	0.3371
Word Order × Cloze scores	10.8719	9.9211	1.096	0.2801
Plausibility × WMC	2.0562	7.1068	0.289	0.7731
Word Order × WMC	6.8922	7.4818	0.921	0.3618
Cloze scores × WMC	0.4233	1.4197	0.298	0.7673
Plausibility × Word Order × Cloze scores	43.7167	16.7118	2.616	0.0090
Plausibility × Word Order × WMC	-9.1708	12.6641	-0.724	0.4691
Plausibility × Cloze scores × WMC	-1.3664	1.2869	-1.062	0.2922
Word Order × Cloze scores × WMC	-0.8687	1.3771	-0.631	0.5316
Plausibility × Word Order × Cloze scores × WMC	1.5516	2.3312	0.666	0.5058

Table A.4 (continued). Linear mixed effects model results for region 3

	Reading time			
	Estimate	SE	*t*	*p*
Random effects structure	(1+Plausibility+Word Order|Subject) + (1+Plausibility+Word Order|Item)			

Table A.5. Linear mixed effects model results for region 4

	Reading time			
	Estimate	SE	*t*	*p*
NS All trials				
(Intercept)	44.38	15.86	2.798	0.0112
Plausibility	-42.55	20.09	-2.118	0.0431
Word Order	12.65	18.89	0.670	0.5089
Plausibility × Word Order	26.23	33.18	0.791	0.4326
Random effects structure	(1+Plausibility*Word Order|Subject) + (1+Plausibility*Word Order|Item)			
NS Correct trials				
(Intercept)	45.14	16.19	2.788	0.0113
Plausibility	-45.93	20.88	-2.200	0.0357
Word Order	11.09	18.48	0.600	0.5533
Plausibility × Word Order	19.83	33.23	0.597	0.5531
Random effects structure	(1+Plausibility*Word Order|Subject) + (1+Plausibility+Word Order|Item)			
L2 All trials				
(Intercept)	-186.12880	30.58524	-6.086	4.91e-07
Plausibility	8.95984	18.62017	0.481	0.633
Word Order	-6.56606	16.09297	-0.408	0.684
Cloze scores	3.38290	5.86257	0.577	0.567
WMC	7.10139	4.21050	1.687	0.100

Table A.5 (continued). Linear mixed effects model results for region 4

	Reading time			
	Estimate	SE	*t*	*p*
Plausibility × Word Order	-28.24919	30.30876	-0.932	0.351
Plausibility × Cloze scores	-1.16996	3.43032	-0.341	0.735
Word Order × Cloze scores	0.98745	2.94077	0.336	0.737
Plausibility × WMC	-3.53936	2.45441	-1.442	0.158
Word Order × WMC	1.02530	2.10593	0.487	0.627
Cloze scores × WMC	-0.67576	0.79153	-0.854	0.399
Plausibility × Word Order × Cloze scores	8.97986	5.83276	1.540	0.124
Plausibility × Word Order × WMC	4.15457	4.18964	0.992	0.322
Plausibility × Cloze scores × WMC	-0.03407	0.46406	-0.073	0.942
Word Order × Cloze scores × WMC	-0.28081	0.40142	-0.700	0.485
Plausibility × Word Order × Cloze scores × WMC	-0.82239	0.79432	-1.035	0.301
Random effects structure	(1+Plausibility+Word Order|Subject) + (1+Plausibility+Word Order+Cloze scores+WMC|Item)			
L2 Correct trials				
(Intercept)	-191.4503	32.6858	-5.857	1.06e-06

Table A.5 (continued). Linear mixed effects model results for region 4

	Reading time			
	Estimate	SE	*t*	*p*
Plausibility	11.9965	23.6344	0.508	0.6151
Word Order	-3.9913	17.3505	-0.230	0.8183
Cloze scores	3.5378	6.2691	0.564	0.5761
WMC	8.3711	4.5452	1.842	0.0734
Plausibility × Word Order	-56.8245	33.8340	-1.680	0.0935
Plausibility × Cloze scores	0.8930	4.3338	0.206	0.8379
Word Order × Cloze scores	0.2187	3.2016	0.068	0.9456
Plausibility × WMC	-4.0741	3.2107	-1.269	0.2116
Word Order × WMC	2.8010	2.4812	1.129	0.2596
Cloze scores × WMC	-1.0006	0.8462	-1.183	0.2449
Plausibility × Word Order × Cloze scores	12.3207	6.3553	1.939	0.0529
Plausibility × Word Order × WMC	3.6983	4.8957	0.755	0.4502
Plausibility × Cloze scores × WMC	-0.2753	0.5971	-0.461	0.6474
Word Order × Cloze scores × WMC	-0.2649	0.4575	-0.579	0.5628
Plausibility × Word Order × Cloze scores × WMC	-0.9871	0.9021	-1.094	0.2742

Table A.5 (continued). Linear mixed effects model results for region 4

	Reading time			
	Estimate	SE	*t*	*p*
Random effects structure	(1+Plausibility*WordOrder|Subject) + (1+Plausibility+Word Order+Cloze scores+WMC|Item),			

Experiment 2

Table A.6. Generalized linear mixed model results, agent identification task

	Offline Accuracy			
	Estimate	SE	*z*	*p*
NS				
(Intercept)	5.1858	0.5904	8.783	<2e-16
Plausibility	0.1880	1.1415	0.165	0.8692
Word Order	-1.1722	0.7032	-1.667	0.0955
Plausibility × Word Order	0.4546	1.4064	0.323	0.7465
Random effects structure	(1+Plausibility|Subject) + (1|Item)			
L2				
(Intercept)	4.36147	0.60317	7.231	4.8e-13
Plausibility	2.55476	0.92434	2.764	0.00571
Word Order	-2.32642	0.91490	-2.543	0.01100
Cloze scores	0.37611	0.12158	3.094	0.00198
Plausibility × Word Order	-3.61710	1.74959	-2.067	0.03870
Plausibility × Cloze scores	0.20104	0.13227	1.520	0.12852
Word Order × Cloze scores	-0.09731	0.13164	-0.739	0.45977
Plausibility × Word Order × Cloze scores	-0.62626	0.24547	-2.551	0.01073
Random effects structure	(1+Plausibility*Word Order|Subject) + (1+Plausibility*Word Order+Cloze scores|Item)			

Table A.7. Generalized linear mixed model results, online comprehension

	Online Accuracy			
	Estimate	SE	*z*	*p*
NS				
(Intercept)	2.93017	0.29344	9.985	< 2e-16
Plausibility	1.47623	0.52390	2.818	0.00484
Word Order	-0.43794	0.48133	-0.910	0.36290
WMC	-0.12848	0.05356	-2.399	0.01644
Plausibility × Word Order	1.94421	0.94267	2.062	0.03917
Plausibility × WMC	-0.15425	0.08890	-1.735	0.08274
Word Order × WMC	-0.03252	0.07652	-0.425	0.67086
Plausibility × Word Order × WMC	-0.17055	0.15084	-1.131	0.25819
Random effects structure	(1+Plausibility*Word Order|Subject) + (1+Plausibility*Word Order+WMC|Item)			
L2				
(Intercept)	0.604258	0.120128	5.030	4.9e-07
Plausibility	1.988215	0.179914	11.051	< 2e-16
Word Order	-0.412219	0.145152	-2.840	0.004512
WMC	-0.021140	0.020014	-1.056	0.290862
Cloze scores	0.107982	0.031353	3.444	0.000573
Plausibility × Word Order	0.171429	0.268963	0.637	0.523884
Plausibility × WMC	-0.019594	0.026972	-0.726	0.467565

Table A.7 (continued). Generalized linear mixed model results, online comprehension

	Online Accuracy			
	Estimate	SE	*z*	*p*
Word Order × WMC	-0.034133	0.022082	-1.546	0.122163
Plausibility × Cloze scores	0.044989	0.042348	1.062	0.288076
Word Order × Cloze scores	0.056460	0.034596	1.632	0.102682
WMC × Cloze scores	-0.005308	0.004346	-1.221	0.221940
Plausibility × Word Order × WMC	0.062500	0.044879	1.393	0.163725
Plausibility × Word Order × Cloze scores	-0.123018	0.070021	-1.757	0.078940
Plausibility × WMC × Cloze scores	0.004148	0.006025	0.688	0.491152
Word Order× WMC × Cloze scores	-0.007707	0.004925	-1.565	0.117587
Plausibility × Word Order × WMC × Cloze scores	-0.011371	0.009942	-1.144	0.252734
Random effects structure	(1+Plausibility*Word Order|Subject) + (1+Plausibility+Word Order+WMC+Cloze scores|Item)			

Table A.8. Linear mixed effects model results for the combined regions of 4 and 5

	Reading time			
	Estimate	SE	*t*	*p*
NS All trials				
(Intercept)	352.484	81.197	4.341	0.000216
Plausibility	-17.496	55.693	-0.314	0.754999
Word Order	269.538	86.558	3.114	0.004877
WMC	3.067	17.176	0.179	0.859880
Plausibility × Word Order	28.326	101.034	0.280	0.779969
Plausibility × WMC	2.125	12.633	0.168	0.867147
Word Order × WMC	22.869	19.094	1.198	0.242395
Plausibility × Word Order × WMC	-49.227	22.158	-2.222	0.029247
Random effects structure	(1+Plausibility*Word Order|Subject) + (1+Plausibility+Word Order+WMC|Item)			
NS Correct trials				
(Intercept)	357.188	82.691	4.320	0.000223
Plausibility	6.590	59.193	0.111	0.911699
Word Order	282.880	89.303	3.168	0.004089
WMC	9.901	17.622	0.562	0.579574
Plausibility × Word Order	30.995	110.335	0.281	0.779154
Plausibility × WMC	-3.363	13.044	-0.258	0.797283
Word Order × WMC	29.423	19.542	1.506	0.144385

Table A.8 (continued). Linear mixed effects model results for the combined regions of 4 and 5

	Reading time			
	Estimate	SE	*t*	*p*
Plausibility × Word Order × WMC	-54.165	24.292	-2.230	0.027201
Random effects structure	(1+Plausibility*Word Order|Subject) + (1+Plausibility+Word Order+WMC|Item)			
L2 All trials				
(Intercept)	676.032	77.148	8.763	1.79e-12
Plausibility	-66.355	54.664	-1.214	0.2367
Word Order	-7.265	53.750	-0.135	0.8932
Cloze scores	14.474	16.805	0.861	0.3943
WMC	-4.291	10.250	-0.419	0.6780
Plausibility × Word Order	18.851	103.086	0.183	0.8549
Plausibility × Cloze scores	10.218	13.746	0.743	0.4628
Word Order × Cloze scores	18.192	13.787	1.320	0.1920
Plausibility × WMC	-13.142	8.641	-1.521	0.1383
Word Order × WMC	18.237	8.653	2.108	0.0392
Cloze scores × WMC	-3.580	2.300	-1.556	0.1285
Plausibility × Word Order × Cloze scores	-48.432	28.740	-1.685	0.0921
Plausibility × Word Order × WMC	20.300	17.052	1.190	0.2340

Table A.8 (continued). Linear mixed effects model results for the combined regions of 4 and 5

	Reading time			
	Estimate	SE	*t*	*p*
Plausibility × Cloze scores × WMC	-1.233	1.921	-0.642	0.5258
Word Order × Cloze scores × WMC	1.458	1.925	0.758	0.4516
Plausibility × Word Order × Cloze scores × WMC	3.416	3.874	0.882	0.3780
Random effects structure	(1+Plausibility+Word Order|Subject) + (1+Plausibility+Word Order+Cloze scores|Item)			
L2 Correct trials				
(Intercept)	667.4981	88.0605	7.580	2.82e-10
Plausibility	-107.0356	93.4931	-1.145	0.259
Word Order	60.5003	90.2891	0.670	0.506
Cloze scores	25.6943	18.4618	1.392	0.171
WMC	-3.2130	11.7129	-0.274	0.785
Plausibility × Word Order	-106.0417	165.5888	-0.640	0.522
Plausibility × Cloze scores	-17.3599	21.7866	-0.797	0.427
Word Order × Cloze scores	10.9006	22.1702	0.492	0.624
Plausibility × WMC	-13.1934	14.6066	-0.903	0.368
Word Order × WMC	22.6373	14.6842	1.542	0.126
Cloze scores × WMC	-2.9795	2.5527	-1.167	0.249

Table A.8 (continued). Linear mixed effects model results for the combined regions of 4 and 5

	Reading time			
	Estimate	SE	*t*	*p*
Plausibility × Word Order × Cloze scores	-17.9331	45.2443	-0.396	0.692
Plausibility × Word Order × WMC	12.9049	28.9100	0.446	0.655
Plausibility × Cloze scores × WMC	0.1035	3.3283	0.031	0.975
Word Order × Cloze scores × WMC	0.2901	3.3330	0.087	0.931
Plausibility × Word Order × Cloze scores × WMC	5.2219	6.6475	0.786	0.432
Random effects structure	(1+Plausibility+Word Order|Subject) + (1+Plausibility+Word Order+Cloze scores+WMC|Item)			

Table A.9. Linear mixed effects model results for region 7

	Reading time			
	Estimate	SE	*t*	*p*
NS All trials				
(Intercept)	96.980	46.363	2.092	0.04753
Plausibility	-118.901	39.603	-3.002	0.00434
Word Order	90.740	48.482	1.872	0.07183
WMC	4.867	9.912	0.491	0.62813
Plausibility × Word Order	-32.197	68.823	-0.468	0.64001
Plausibility × WMC	-8.845	8.354	-1.059	0.29393
Word Order × WMC	5.357	10.419	0.514	0.61112
Plausibility × Word Order × WMC	17.933	15.145	1.184	0.23691
Random effects structure	(1+Plausibility+Word Order|Subject) + (1+Plausibility+Word Order+WMC|Item)			
NS Correct trials				
(Intercept)	96.195	50.503	1.905	0.0687
Plausibility	-106.722	43.832	-2.435	0.0182
Word Order	88.265	55.603	1.587	0.1240
WMC	5.024	10.647	0.472	0.6415
Plausibility × Word Order	-13.105	77.354	-0.169	0.8655
Plausibility × WMC	-7.428	9.341	-0.795	0.4291
Word Order × WMC	7.548	12.141	0.622	0.5391

Table A.9 (continued). Linear mixed effects model results for region 7

	Reading time			
	Estimate	SE	*t*	*p*
Plausibility × Word Order × WMC	19.854	16.913	1.174	0.2408
Random effects structure	(1+Plausibility+Word Order|Subject) + (1+Plausibility+Word Order+WMC||Item)			
L2 All trials				
(Intercept)	226.8960	50.1666	4.523	3e-05
Plausibility	-134.6775	54.2833	-2.481	0.0179
Word Order	39.3021	39.8834	0.985	0.3297
Cloze scores	15.0138	11.0666	1.357	0.1833
WMC	-6.6251	6.9703	-0.950	0.3482
Plausibility × Word Order	82.6183	66.3336	1.245	0.2131
Plausibility × Cloze scores	-33.3554	14.5350	-2.295	0.0275
Word Order × Cloze scores	8.1455	9.7905	0.832	0.4078
Plausibility × WMC	7.1046	9.1638	0.775	0.4431
Word Order × WMC	-6.6993	6.2374	-1.074	0.2858
Cloze scores × WMC	-1.5780	1.5643	-1.009	0.3198
Plausibility × Word Order × Cloze scores	12.4620	18.8691	0.660	0.5091
Plausibility × Word Order × WMC	-1.1530	11.2809	-0.102	0.9186

Table A.9 (continued). Linear mixed effects model results for region 7

	Reading time			
	Estimate	SE	*t*	*p*
Plausibility × Cloze scores × WMC	-0.2815	2.0450	-0.138	0.8913
Word Order × Cloze scores × WMC	-0.5178	1.3815	-0.375	0.7088
Plausibility × Word Order × Cloze scores × WMC	-0.1550	2.5237	-0.061	0.9510
Random effects structure	(1+Plausibility+Word Order|Subject) + (1+Plausibility+Word Order+Cloze scores+WMC||Item)			
L2 Correct trials				
(Intercept)	213.49119	53.49196	3.991	0.000195
Plausibility	-151.48321	77.45316	-1.956	0.057176
Word Order	67.48416	60.45820	1.116	0.267787
Cloze scores	21.05485	12.40785	1.697	0.097158
WMC	-7.02803	7.86986	-0.893	0.376740
Plausibility × Word Order	33.37747	104.17950	0.320	0.748748
Plausibility × Cloze scores	-52.49082	19.97123	-2.628	0.012043
Word Order × Cloze scores	17.11874	14.67668	1.166	0.245619
Plausibility × WMC	4.28476	12.81390	0.334	0.739657
Word Order × WMC	-13.42802	9.69934	-1.384	0.168136
Cloze scores × WMC	-2.02543	1.79052	-1.131	0.263958

Table A.9 (continued). Linear mixed effects model results for region 7

	Reading time			
	Estimate	SE	*t*	*p*
Plausibility × Word Order × Cloze scores	-19.47566	28.63469	-0.680	0.496600
Plausibility × Word Order × WMC	-7.38637	18.44209	-0.401	0.688867
Plausibility × Cloze scores × WMC	0.74914	2.94065	0.255	0.800055
Word Order × Cloze scores × WMC	0.22843	2.22809	0.103	0.918478
Plausibility × Word Order × Cloze scores × WMC	-0.03837	4.23815	-0.009	0.992778
Random effects structure	(1+Plausibility+Word Order|Subject) + (1+Plausibility+Word Order+Cloze scores+WMC|Item)			

Table A.10. Linear mixed effects model results for region 8

	Reading time			
	Estimate	SE	*t*	*p*
NS All trials				
(Intercept)	-44.0094	28.5182	-1.543	0.136531
Plausibility	-57.0868	16.3175	-3.499	0.000941
Word Order	18.3891	15.7406	1.168	0.247315
WMC	7.2617	6.2360	1.164	0.255818
Plausibility × Word Order	10.9588	27.2827	0.402	0.688007
Plausibility × WMC	-1.7896	2.9970	-0.597	0.550831
Word Order × WMC	0.2699	3.0947	0.087	0.930649
Plausibility × Word Order × WMC	4.4555	6.1629	0.723	0.469987
Random effects structure	(1+Plausibility+Word Order|Subject) + (1+Plausibility+Word Order+WMC|Item)			
NS Correct trials				
(Intercept)	-50.7117	28.1263	-1.803	0.08443
Plausibility	-53.4645	17.7859	-3.006	0.00408
Word Order	9.3077	14.6664	0.635	0.52602
WMC	7.5085	6.1438	1.222	0.23361
Plausibility × Word Order	21.5162	30.0555	0.716	0.47610
Plausibility × WMC	-3.3123	3.2190	-1.029	0.30522
Word Order × WMC	-0.8368	3.1071	-0.269	0.78784

Table A.10 (continued). Linear mixed effects model results for region 8

	Reading time			
	Estimate	SE	*t*	*p*
Plausibility × Word Order × WMC	5.8648	6.5400	0.897	0.37156
Random effects structure	(1+Plausibility*Word Order|Subject) + (1+Plausibility*Word Order+WMC|Item)			
L2 All trials				
(Intercept)	-214.50779	25.60030	-8.379	1.9e-11
Plausibility	-41.05095	18.52184	-2.216	0.0326
Word Order	26.15704	17.61572	1.485	0.1408
Cloze scores	-8.42429	5.90475	-1.427	0.1620
WMC	4.12928	3.69230	1.118	0.2707
Plausibility × Word Order	-5.64876	33.48245	-0.169	0.8661
Plausibility × Cloze scores	2.38531	4.77544	0.499	0.6200
Word Order × Cloze scores	-11.34478	4.69394	-2.417	0.0174
Plausibility × WMC	-1.58773	3.02613	-0.525	0.6025
Word Order × WMC	1.54718	2.96496	0.522	0.6029
Cloze scores × WMC	0.86723	0.82403	1.052	0.2996
Plausibility × Word Order × Cloze scores	-3.87861	9.33164	-0.416	0.6779
Plausibility × Word Order × WMC	1.88491	5.61029	0.336	0.7371

Table A.10 (continued). Linear mixed effects model results for region 8

	Reading time			
	Estimate	SE	*t*	*p*
Plausibility × Cloze scores × WMC	0.02766	0.67110	0.041	0.9673
Word Order × Cloze scores × WMC	0.62343	0.65809	0.947	0.3457
Plausibility × Word Order × Cloze scores × WMC	0.63697	1.25808	0.506	0.6130
Random effects structure	(1+Plausibility*Word Order|Subject) + (1+Plausibility+Word Order+Cloze scores+WMC|Item)			
L2 Correct trials				
(Intercept)	-212.9135	27.7002	-7.686	7.08e-10
Plausibility	-62.1115	36.5484	-1.699	0.0987
Word Order	27.3859	26.9395	1.017	0.3113
Cloze scores	-12.1072	6.7852	-1.784	0.0823
WMC	4.1932	4.2619	0.984	0.3312
Plausibility × Word Order	8.5797	52.3171	0.164	0.8698
Plausibility × Cloze scores	3.1488	9.4468	0.333	0.7411
Word Order × Cloze scores	-15.9305	7.0680	-2.254	0.0257
Plausibility × WMC	-0.4314	6.0795	-0.071	0.9438
Word Order × WMC	5.2588	4.6303	1.136	0.2574
Cloze scores × WMC	0.8805	0.9657	0.912	0.3674

Table A.10 (continued). Linear mixed effects model results for region 8

	Reading time			
	Estimate	SE	*t*	*p*
Plausibility × Word Order × Cloze scores	-11.5951	14.1905	-0.817	0.4143
Plausibility × Word Order × WMC	-3.7329	8.9685	-0.416	0.6774
Plausibility × Cloze scores × WMC	-0.4792	1.3982	-0.343	0.7338
Word Order × Cloze scores × WMC	0.9259	1.0695	0.866	0.3878
Plausibility × Word Order × Cloze scores × WMC	-0.2945	2.0783	-0.142	0.8874
Random effects structure	(1+Plausibility*Word Order|Subject) + (1+Plausibility+Word Order+Cloze scores+WMC|Item)			

Experiment 3

Table A.11. Generalized linear mixed model results, acceptability judgment task

	Ratings			
	Estimate	SE	*z*	*p*
NS				
(Intercept)	-4.419e-16	1.637e-02	0.000	1.000000
Information Structure	-9.256e-01	1.713e-01	-5.403	3.4e-05
Word Order	-6.185e-01	1.333e-01	-4.641	0.000186
WMC	-2.942e-17	3.533e-03	0.000	1.000000
Information Structure × Word Order	-1.782e-01	6.216e-02	-2.866	0.004208
Information Structure × WMC	-1.995e-03	3.861e-02	-0.052	0.959356
Word Order × WMC	2.506e-02	3.017e-02	0.831	0.416816
Information Structure × Word Order × WMC	3.799e-02	1.438e-02	2.642	0.008385
Random effects structure	(1+Information Structure+Word Order||Subject) + (1+Information Structure+Word Order||Item)			
L2				
(Intercept)	-3.843e-18	1.604e-02	0.000	1.0000
Information Structure	-3.205e-01	1.465e-01	-2.187	0.0439
Word Order	-1.059e+00	1.449e-01	-7.311	1.17e-06
WMC	-4.267e-18	2.594e-03	0.000	1.0000
Cloze scores	1.552e-17	4.147e-03	0.000	1.0000
Information Structure × Word Order	-5.261e-02	6.344e-02	-0.829	0.4071

Table A.11 (continued). Generalized linear mixed model results, acceptability judgment task

	Ratings			
	Estimate	SE	*z*	*p*
Information Structure × WMC	1.839e-03	2.398e-02	0.077	0.9398
Word Order × WMC	7.564e-03	2.330e-02	0.325	0.7497
Information Structure × Cloze scores	3.042e-03	3.832e-02	0.079	0.9377
Word Order × Cloze scores	-7.731e-02	3.724e-02	-2.076	0.0544
WMC × Cloze scores	-1.183e-19	6.710e-04	0.000	1.0000
Information Structure × Word Order × WMC	8.081e-03	1.038e-02	0.779	0.4363
Information Structure × Word Order × Cloze scores	-9.118e-03	1.659e-02	-0.550	0.5826
Information Structure × WMC × Cloze scores	-6.099e-03	6.201e-03	-0.984	0.3400
Word Order × WMC × Cloze scores	8.009e-03	6.026e-03	1.329	0.2025
Information Structure × Word Order × WMC × Cloze scores	2.320e-03	2.684e-03	0.864	0.3876
Random effects structure	(1+Information Structure+Word Order||Subject) + (1+Word Order+WMC||Item)			

Table A.12. Generalized linear mixed model results, online comprehension

	Online accuracy			
	Estimate	SE	*z*	*p*
NS				
(Intercept)	8.02009	0.89083	9.003	< 2e-16
Information Structure	-1.50655	1.68588	-0.894	0.3715
Word Order	0.93143	1.71231	0.544	0.5865
WMC	-0.12603	0.06704	-1.880	0.0601
Information Structure × Word Order	-13.23055	2.25933	-5.856	4.74e-09
Information Structure × WMC	0.19402	0.25209	0.770	0.4415
Word Order × WMC	-0.03527	0.25396	-0.139	0.8895
Information Structure × Word Order × WMC	0.67162	0.27356	2.455	0.0141
Random effects structure	(1+Information Structure+Word Order|Subject) + (1+Information Structure+Word Order|Item)			
L2				
(Intercept)	0.8814948	0.2093678	4.210	2.55e-05
Information Structure	0.2509729	0.1549932	1.619	0.1054
Word Order	-0.7477326	0.1563480	-4.782	1.73e-06
Cloze scores	0.0294519	0.0442151	0.666	0.5053
WMC	-0.0516197	0.0281882	-1.831	0.0671
Information Structure × Word Order	-0.5482811	0.3103082	-1.767	0.0772
Information Structure × Cloze scores	0.0047259	0.0422537	0.112	0.9109

Table A.12 (continued). Generalized linear mixed model results, online comprehension

	Online accuracy			
	Estimate	SE	*z*	*p*
Word Order × Cloze scores	0.0408095	0.0417550	0.977	0.3284
Information Structure × WMC	0.0270515	0.0279634	0.967	0.3333
Word Order × WMC	-0.0009057	0.0281454	-0.032	0.9743
Cloze scores × WMC	-0.0155917	0.0073081	-2.133	0.0329
Information Structure × Word Order × Cloze scores	-0.0660641	0.0838730	-0.788	0.4309
Information Structure × Word Order × WMC	0.0452448	0.0554473	0.816	0.4145
Information Structure × Cloze scores × WMC	0.0046699	0.0073205	0.638	0.5235
Word Order × Cloze scores × WMC	-0.0003970	0.0071911	-0.055	0.9560
Information Structure × Word Order × Cloze scores × WMC	-0.0007506	0.0144005	-0.052	0.9584
Random effects structure	(1|Subject) + (1|Item)			

Table A.13. Linear mixed effects model results for region 4

	Reading time			
	Estimate	SE	*t*	*p*
NS All trials				
(Intercept)	-30.8957	7.5995	-4.066	0.000669
Information Structure	-2.8984	7.3601	-0.394	0.696426
Word Order	13.0553	7.2949	1.790	0.080026
WMC	0.9873	1.7011	0.580	0.568843
Information Structure × Word Order	16.0013	13.7919	1.160	0.252558
Information Structure × WMC	0.3848	1.5901	0.242	0.809478
Word Order × WMC	1.9104	1.5941	1.198	0.234953
Information Structure × Word Order × WMC	-3.8965	3.2709	-1.191	0.239533
Random effects structure	(1+Information Structure*Word Order|Subject) + (1+Information Structure+Word Order+ WMC|Item)			
NS Correct trials				
(Intercept)	-30.6523	7.4187	-4.132	0.000554
Information Structure	-4.2449	7.7126	-0.550	0.586231
Word Order	13.3273	7.3972	1.802	0.078675
WMC	1.1489	1.6511	0.696	0.495391
Information Structure × Word Order	16.1354	14.0338	1.150	0.257665
Information Structure × WMC	0.7791	1.6433	0.474	0.637360

Table A.13 (continued). Linear mixed effects model results for region 4

	Reading time			
	Estimate	SE	*t*	*p*
Word Order × WMC	2.6014	1.6132	1.613	0.112348
Information Structure × Word Order × WMC	-3.4225	3.3432	-1.024	0.311784
Random effects structure	(1+Information Structure*Word Order|Subject) + (1+Information Structure+Word Order+ WMC|Item)			
L2 All trials				
(Intercept)	23.8359	25.7321	0.926	0.3630
Information Structure	-187.4467	31.4105	-5.968	7.01e-06
Word Order	63.1595	30.4621	2.073	0.0446
Cloze scores	1.3013	6.3285	0.206	0.8390
WMC	0.5380	3.6266	0.148	0.8839
Information Structure × Word Order	-9.1027	49.3830	-0.184	0.8538
Information Structure × Cloze scores	7.2788	7.4421	0.978	0.3425
Word Order × Cloze scores	1.7866	6.8734	0.260	0.7960
Information Structure × WMC	-3.7860	4.6314	-0.817	0.4257
Word Order × WMC	-2.8710	4.3450	-0.661	0.5118
Cloze scores × WMC	-0.5011	0.9383	-0.534	0.6007
Information Structure × Word Order × Cloze scores	18.4131	12.9653	1.420	0.1559

Table A.13 (continued). Linear mixed effects model results for region 4

	Reading time			
	Estimate	SE	*t*	*p*
Information Structure × Word Order × WMC	-8.5781	8.1345	-1.055	0.2920
Information Structure × Cloze scores × WMC	0.4850	1.2189	0.398	0.6958
Word Order × Cloze scores × WMC	-1.7290	1.1119	-1.555	0.1266
Information Structure × Word Order × Cloze scores × WMC	1.0837	2.1207	0.511	0.6095
Random effects structure	(1+Information Structure+Word Order|Subject) + (1+Information Structure+Word Order+ Cloze scores|Item)			
L2 Correct trials				
(Intercept)	23.2725	27.0652	0.860	0.399333
Information Structure	-171.4174	39.2557	-4.367	0.000318
Word Order	50.6131	35.0262	1.445	0.154912
Cloze scores	3.9965	6.5928	0.606	0.552101
WMC	-0.7913	3.8563	-0.205	0.840320
Information Structure × Word Order	37.4163	61.1686	0.612	0.541000
Information Structure × Cloze scores	-4.3899	9.9478	-0.441	0.664331
Word Order × Cloze scores	9.0499	7.9641	1.136	0.257992
Information Structure × WMC	-2.7999	6.0225	-0.465	0.648368

Table A.13 (continued). Linear mixed effects model results for region 4

	Reading time			
	Estimate	SE	*t*	*p*
Word Order × WMC	-3.4345	4.7955	-0.716	0.475363
Cloze scores × WMC	-0.7925	1.0000	-0.793	0.440922
Information Structure × Word Order × Cloze scores	-11.2363	15.7814	-0.712	0.476768
Information Structure × Word Order × WMC	-13.2393	9.4457	-1.402	0.161597
Information Structure × Cloze scores × WMC	-0.6885	1.5800	-0.436	0.668668
Word Order × Cloze scores × WMC	-1.8769	1.2408	-1.513	0.133204
Information Structure × Word Order × Cloze scores × WMC	-2.3634	2.4859	-0.951	0.342184
Random effects structure	(1+Information Structure+Word Order|Subject) + (1+Information Structure+Word Order+ Cloze scores|Item)			

Table A.14. Linear mixed effects model results for region 5

	Reading time			
	Estimate	SE	*t*	*p*
NS All trials				
(Intercept)	-23.2953	6.9884	-3.333	0.00359
Information Structure	4.4304	10.8014	0.410	0.68542
Word Order	21.0043	9.8726	2.128	0.04352
WMC	0.5033	1.5593	0.323	0.75064
Information Structure × Word Order	37.5172	15.1796	2.472	0.01371
Information Structure × WMC	1.0288	2.1860	0.471	0.64297
Word Order × WMC	-1.1919	2.2808	-0.523	0.60557
Information Structure × Word Order × WMC	-2.5375	3.6238	-0.700	0.48421
Random effects structure	(1+Information Structure+Word Order|Subject) + (1+Information Structure+Word Order + WMC|Item)			
NS Correct trials				
(Intercept)	-22.7750	7.0372	-3.236	0.00433
Information Structure	2.9999	11.1915	0.268	0.79087
Word Order	20.4999	10.2089	2.008	0.05610
WMC	0.6516	1.5469	0.421	0.67869
Information Structure × Word Order	36.9431	15.2551	2.422	0.01568
Information Structure × WMC	1.1290	2.1637	0.522	0.60743

Table A.14 (continued). Linear mixed effects model results for region 5

	Reading time			
	Estimate	SE	*t*	*p*
Word Order × WMC	-0.7963	2.3727	-0.336	0.73984
Information Structure × Word Order × WMC	-2.8794	3.6640	-0.786	0.43244
Random effects structure	(1+Information Structure+Word Order|Subject) + (1+Information Structure+Word Order|Item)			
L2 All trials				
(Intercept)	252.9697	36.3302	6.963	8.11e-08
Information Structure	-76.0414	32.2011	-2.361	0.029826
Word Order	129.0158	30.1568	4.278	0.000115
Cloze scores	-17.5752	7.9793	-2.203	0.040903
WMC	-2.8270	4.9196	-0.575	0.572998
Information Structure × Word Order	76.3030	54.9889	1.388	0.165665
Information Structure × Cloze scores	-5.2206	7.5390	-0.692	0.499281
Word Order × Cloze scores	-14.7055	7.4460	-1.975	0.051322
Information Structure × WMC	-1.5843	4.8200	-0.329	0.746702
Word Order × WMC	-0.5173	4.7291	-0.109	0.913121
Cloze scores × WMC	-0.7294	1.2459	-0.585	0.566442
Information Structure × Word Order × Cloze scores	-12.0328	14.4773	-0.831	0.406139

Table A.14 (continued). Linear mixed effects model results for region 5

	Reading time			
	Estimate	SE	*t*	*p*
Information Structure × Word Order × WMC	-6.0822	9.0857	-0.669	0.503414
Information Structure × Cloze scores × WMC	-1.9764	1.2411	-1.592	0.131653
Word Order × Cloze scores × WMC	-1.6751	1.2050	-1.390	0.167883
Information Structure × Word Order × Cloze scores × WMC	-2.3794	2.3464	-1.014	0.310883
Random effects structure	(1+Information Structure+Word Order|Subject) + (1+Information Structure+Word Order+ Cloze scores+WMC|Item)			
L2 Correct trials				
(Intercept)	239.686	39.284	6.101	7.17e-07
Information Structure	-78.756	38.265	-2.058	0.04349
Word Order	128.416	36.260	3.542	0.00076
Cloze scores	-15.955	8.080	-1.975	0.06557
WMC	-3.814	4.856	-0.786	0.44475
Information Structure × Word Order	80.705	69.619	1.159	0.24690
Information Structure × Cloze scores	-4.144	9.126	-0.454	0.65029
Word Order × Cloze scores	-18.110	9.263	-1.955	0.05492
Information Structure × WMC	1.382	5.502	0.251	0.80203

Table A.14 (continued). Linear mixed effects model results for region 5

	Reading time			
	Estimate	SE	*t*	*p*
Word Order × WMC	-2.229	5.578	-0.400	0.69088
Cloze scores × WMC	-1.088	1.259	-0.865	0.40107
Information Structure × Word Order × Cloze scores	-10.964	17.757	-0.617	0.53721
Information Structure × Word Order × WMC	-8.182	10.570	-0.774	0.43920
Information Structure × Cloze scores × WMC	-2.450	1.434	-1.708	0.08949
Word Order × Cloze scores × WMC	-1.885	1.441	-1.309	0.19607
Information Structure × Word Order × Cloze scores × WMC	-2.891	2.771	-1.043	0.29720
Random effects structure	(1+Information Structure+Word Order|Subject) + (1+Information Structure+Word Order+ Cloze scores+WMC||Item)			

Table A.15. Linear mixed effects model results for region 6

	Reading time			
	Estimate	SE	*t*	*p*
NS All trials				
(Intercept)	-22.5427	6.0459	-3.729	0.00092
Information Structure	19.8585	9.6913	2.049	0.05188
Word Order	11.7429	9.5373	1.231	0.22740
WMC	2.1902	1.2552	1.745	0.09483
Information Structure × Word Order	53.4098	18.2835	2.921	0.00674
Information Structure × WMC	2.1826	2.2467	0.971	0.34061
Word Order × WMC	0.8705	2.0984	0.415	0.68093
Information Structure × Word Order × WMC	2.1980	4.4103	0.498	0.62133
Random effects structure	(1+Information Structure*Word Order|Subject) + (1+Information Structure+Word Order+ WMC|Item)			
NS Correct trials				
(Intercept)	-22.715	6.046	-3.757	0.000845
Information Structure	17.279	9.844	1.755	0.092340
Word Order	11.909	9.914	1.201	0.239348
WMC	2.310	1.244	1.857	0.076848
Information Structure × Word Order	53.749	18.764	2.865	0.007883
Information Structure × WMC	2.568	2.287	1.123	0.272120

Table A.15 (continued). Linear mixed effects model results for region 6

	Reading time			
	Estimate	SE	*t*	*p*
Word Order × WMC	1.185	2.159	0.549	0.587119
Information Structure × Word Order × WMC	1.957	4.517	0.433	0.667649
Random effects structure	(1+Information Structure*Word Order|Subject) + (1+Information Structure+Word Order+WMC|Item)			
L2 All trials				
(Intercept)	456.045	59.559	7.657	1.47e-08
Information Structure	-5.058	45.831	-0.110	0.9132
Word Order	135.827	55.253	2.458	0.0200
Cloze scores	-31.554	13.942	-2.263	0.0342
WMC	-6.679	8.146	-0.820	0.4238
Information Structure × Word Order	57.733	83.306	0.693	0.4885
Information Structure × Cloze scores	4.843	11.406	0.425	0.6758
Word Order × Cloze scores	-19.625	13.195	-1.487	0.1484
Information Structure × WMC	11.748	7.148	1.644	0.1162
Word Order × WMC	-1.746	8.325	-0.210	0.8354
Cloze scores × WMC	-2.733	2.080	-1.314	0.2073
Information Structure × Word Order × Cloze scores	-28.297	22.063	-1.283	0.2000

Table A.15 (continued). Linear mixed effects model results for region 6

	Reading time			
	Estimate	SE	*t*	*p*
Information Structure × Word Order × WMC	-11.485	13.721	-0.837	0.4028
Information Structure × Cloze scores × WMC	1.840	1.881	0.978	0.3396
Word Order × Cloze scores × WMC	-4.289	2.130	-2.014	0.0541
Information Structure × Word Order × Cloze scores × WMC	-6.876	3.570	-1.926	0.0544
Random effects structure	(1+Information Structure+Word Order|Subject) + (1+Information Structure+Word Order+ Cloze scores+WMC|Item)			
L2 Correct trials				
(Intercept)	444.1144	68.7669	6.458	9.55e-07
Information Structure	-35.8176	55.0663	-0.650	0.5175
Word Order	117.8861	77.1370	1.528	0.1401
Cloze scores	-28.1520	16.6777	-1.688	0.1073
WMC	-5.4199	9.7876	-0.554	0.5877
Information Structure × Word Order	3.4659	100.7880	0.034	0.9726
Information Structure × Cloze scores	12.2123	13.4088	0.911	0.3629
Word Order × Cloze scores	-10.1434	18.4104	-0.551	0.5882
Information Structure × WMC	13.5067	7.8019	1.731	0.0842

Table A.15 (continued). Linear mixed effects model results for region 6

	Reading time			
	Estimate	SE	*t*	*p*
Word Order × WMC	0.6956	11.2079	0.062	0.9512
Cloze scores × WMC	-2.0166	2.5228	-0.799	0.4364
Information Structure × Word Order × Cloze scores	-4.9668	26.3035	-0.189	0.8503
Information Structure × Word Order × WMC	-9.7085	15.5492	-0.624	0.5326
Information Structure × Cloze scores × WMC	1.5962	2.1076	0.757	0.4493
Word Order × Cloze scores × WMC	-2.1656	2.9141	-0.743	0.4674
Information Structure × Word Order × Cloze scores × WMC	-9.6547	4.0699	-2.372	0.0180
Random effects structure	(1+Information Structure+Word Order|Subject) + (1+Information Structure+Word Order+ Cloze scores+WMC|Item)			